

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 15, 2005, 15:01:47 ; Search time 2260 Seconds  
(without alignments)  
9332.392 Million cell updates/sec

Title: US-10-009-579-5\_COPY\_3115\_3560  
Perfect score: 446  
Sequence: 1 gtaatggcagatctctgct.....ctggaaggtctctgctgt 446

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 4526729 seqs, 23644849745 residues  
Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hg: \*  
3: gb\_in: \*  
4: gb\_on: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Query Match		Length DB		ID		Description	
Result No.	Score	Match	Length	DB	ID	Score	Match	Length	DB	ID	Score	Match
1	446	100.0	4069	9	AY148099	446	100.0	4069	9	AY148099	446	100.0
2	446	100.0	4282	6	AX254778	446	100.0	4282	6	AX254778	446	100.0
3	446	100.0	171987	9	AC079775	446	100.0	171987	9	AC079775	446	100.0
C 4	173.4	38.9	97288	9	AL356055	173.4	38.9	97288	9	AL356055	173.4	38.9
C 5	172.2	38.6	110130	9	AL360268	172.2	38.6	110130	9	AL360268	172.2	38.6
C 6	171.6	38.5	58723	9	AL391237	171.6	38.5	58723	9	AL391237	171.6	38.5
C 7	171.4	38.4	158405	9	AC021763	171.4	38.4	158405	9	AC021763	171.4	38.4
8	171.4	38.4	187200	2	AP001379	171.4	38.4	187200	2	AP001379	171.4	38.4
9	171	38.3	59829	9	AC064841	171	38.3	59829	9	AC064841	171	38.3
C 10	171	38.3	178670	2	AC009290	171	38.3	178670	2	AC009290	171	38.3
C 11	170.6	38.3	176075	5	AC017079	170.6	38.3	176075	5	AC017079	170.6	38.3
12	170	38.1	568	6	AX871940	170	38.1	568	6	AX871940	170	38.1
C 13	170	38.1	568	6	BD152002	170	38.1	568	6	BD152002	170	38.1
C 14	170	38.1	163515	9	AC144522	170	38.1	163515	9	AC144522	170	38.1
C 15	170	38.1	185593	2	AC026856	170	38.1	185593	2	AC026856	170	38.1
C 16	169.6	38.0	66266	2	AP000573	169.6	38.0	66266	2	AP000573	169.6	38.0
C 17	169.6	38.0	77265	2	AC099830	169.6	38.0	77265	2	AC099830	169.6	38.0
C 18	169.6	38.0	166356	9	AP001453	169.6	38.0	166356	9	AP001453	169.6	38.0
C 19	169.6	38.0	174562	2	AC005848	169.6	38.0	174562	2	AC005848	169.6	38.0

20	169.2	37.9	3252	6	AX833480	169.2	37.9	3252	6	AX833480
21	169.2	37.9	3252	9	AK095253	169.2	37.9	3252	9	AK095253
C 22	169.2	37.9	67374	9	CR394533	169.2	37.9	67374	9	CR394533
C 23	169.2	37.9	139255	9	AL512642	169.2	37.9	139255	9	AL512642
C 24	169.2	37.9	175320	9	AC009283	169.2	37.9	175320	9	AC009283
C 25	169	37.9	85123	9	AC125616	169	37.9	85123	9	AC125616
C 26	169	37.9	91672	9	AC006468	169	37.9	91672	9	AC006468
C 27	169	37.9	137242	9	AC063943	169	37.9	137242	9	AC063943
C 28	169	37.9	146128	9	AC026358	169	37.9	146128	9	AC026358
C 29	169	37.9	203050	2	HS44N10	169	37.9	203050	2	HS44N10
C 30	168.4	37.8	88119	9	AL450325	168.4	37.8	88119	9	AL450325
C 31	168.4	37.8	157268	2	AC025063	168.4	37.8	157268	2	AC025063
C 32	168.4	37.8	175067	2	AC040898	168.4	37.8	175067	2	AC040898
C 33	168.4	37.8	230039	2	AC090391	168.4	37.8	230039	2	AC090391
C 34	168.2	37.7	148285	2	AL451053	168.2	37.7	148285	2	AL451053
C 35	168.2	37.7	192096	9	AL590133	168.2	37.7	192096	9	AL590133
C 36	167.8	37.6	68470	2	AC116011	167.8	37.6	68470	2	AC116011
C 37	167.8	37.6	209512	9	CNS0180V	167.8	37.6	209512	9	CNS0180V
C 38	167.8	37.6	213648	9	AL158040	167.8	37.6	213648	9	AL158040
C 39	167.6	37.6	156203	2	AP001386	167.6	37.6	156203	2	AP001386
C 40	167.6	37.6	170172	2	AC104765	167.6	37.6	170172	2	AC104765
C 41	167.6	37.6	183055	9	AC120349	167.6	37.6	183055	9	AC120349
C 42	167.4	37.5	84001	9	AL162739	167.4	37.5	84001	9	AL162739
C 43	167.4	37.5	110459	9	AC117378	167.4	37.5	110459	9	AC117378
C 44	166.8	37.4	147124	9	AC005913	166.8	37.4	147124	9	AC005913
C 45	166.8	37.4	200070	9	AC008746	166.8	37.4	200070	9	AC008746

ALIGNMENTS

RESULT 1  
AY148099  
LOCUS Homo sapiens EGP2 (TACSTD1) gene, promoter region and 5'UTR. PRI 12-NOV-2002  
DEFINITION Homo sapiens EGP2 (TACSTD1) gene, promoter region and 5'UTR.  
ACCESSION AY148099  
VERSION AY148099.1 GI:24935271  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 4069)  
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS McLaughlin,P.M.J., Trzpis,M., Kroesen,B.-J., Helfrich,W.,  
Terpstra,P., Ruiters,M.H.C., de Leij,L.F.M.H. and Harmsen,M.C.  
TITLE Use of the EGP-2 promoter for targeted expression of heterologous  
genes in carcinoma-derived cell lines  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4069)  
McLaughlin,P.M.J., Trzpis,M., Kroesen,B.-J., Helfrich,W.,  
Terpstra,P., Ruiters,M.H.C., de Leij,L.F.M.H. and Harmsen,M.C.  
AUTHORS Direct Submission  
TITLE Submitted (04-SEP-2002) Path-LabMed-Mol-Biol, University Groningen,  
JOURNAL Hanzeplein 1, Groningen 9713GZ, The Netherlands  
FEATURES  
source  
1..4069  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2p16.3"  
1..>4069  
/gene="TACSTD1"  
/notes="synonyms: EGP2, GA733-2"  
1..3974  
/gene="TACSTD1"  
3975..>4069  
/gene="TACSTD1"  
/product="EGP2"  
/note="putative transcription start"  
3975..>4069  
/gene="TACSTD1"  
5'UTR  
ORIGIN

```
Query Match 100.0%; Score 446; DB 9; Length 4069;
Best Local Similarity 100.0%; Pred. No. 3.5e-96;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTAATGCGCAGTCTCTGCTCACTGCAACCTCAGCCTCCCGCTAGCTGGATTACAGGC 60
DB 3112 GTAATGCGCAGTCTCTGCTCACTGCAACCTCAGCCTCCCGCTAGCTGGATTACAGGC 3171
QY 61 ATCGGCCACACCCCGGGCTAAATTTGTATCTTTTAGTAGAGAGCGGCTTCTCCATGTT 120
DB 3172 ATCGGCCACACCCCGGGCTAAATTTGTATCTTTTAGTAGAGAGCGGCTTCTCCATGTT 3231
QY 121 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCGCCCGCTCGGCTCCCAAGT 180
DB 3232 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCGCCCGCTCGGCTCCCAAGT 3291
QY 181 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCGCCCGCTCGGCTCCCAAGT 3300
DB 3292 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCGCCCGCTCGGCTCCCAAGT 3351
QY 241 AGTGTCTAGAAATGCTTATGAAAACGAAAAAGAAATTAATTAAGAGTAATTAATAAGAAAC 300
DB 3352 AGTGTCTAGAAATGCTTATGAAAACGAAAAAGAAATTAATTAAGAGTAATTAATAAGAAAC 3411
QY 301 ACTCATTTCTCCAGAGAGCAAGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
DB 3412 ACTCATTTCTCCAGAGAGCAAGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3471
QY 361 TTTCTAATTTCAAGAGATATAATTAATTAAGAGTAATTAATAAGAGTAATTAATAAGAGTAAT 420
DB 3472 TTTCTAATTTCAAGAGATATAATTAATTAAGAGTAATTAATAAGAGTAATTAATAAGAGTAAT 3531
QY 421 AGTGTCTGGAAGGTTCTCTGCTGT 446
DB 3532 AGTGTCTGGAAGGTTCTCTGCTGT 3557

RESULT 2
AX254778 4282 bp DNA linear PAT 10-OCT-2001
LOCUS Sequence 5 from Patent W00171015.
DEFINITION AX254778
ACCESSION AX254778
VERSION AX254778.1 GI:16074440
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS de Leij,L.F., Ruiter,M.H., McLaughlin,P.M., Harmsen,M.C., van der
Molen,H., Terpstra,P. and Dokter,W.H.
TITLE Non-squamous epithelium-specific transcription
JOURNAL Patent: WO 0171015-A 5 27-SEP-2001;
Rijksuniversiteit Groningen (NL)
FEATURES
source
1. .4282
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
misc_feature 1. .4282
/note="EGP-2 promoter sequence from -3967 to +315"
ORIGIN
Query Match 100.0%; Score 446; DB 6; Length 4282;
Best Local Similarity 100.0%; Pred. No. 3.5e-96;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTAATGCGCAGTCTCTGCTCACTGCAACCTCAGCCTCCCGCTAGCTGGATTACAGGC 60
DB 3115 GTAATGCGCAGTCTCTGCTCACTGCAACCTCAGCCTCCCGCTAGCTGGATTACAGGC 3174
QY 61 ATCGGCCACACCCCGGGCTAAATTTGTATCTTTTAGTAGAGAGCGGCTTCTCCATGTT 120
DB 3172 ATCGGCCACACCCCGGGCTAAATTTGTATCTTTTAGTAGAGAGCGGCTTCTCCATGTT 3231
QY 121 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCGCCCGCTCGGCTCCCAAGT 180
DB 3232 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCGCCCGCTCGGCTCCCAAGT 3291
QY 181 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCGCCCGCTCGGCTCCCAAGT 3300
DB 3292 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCGCCCGCTCGGCTCCCAAGT 3351
QY 241 AGTGTCTAGAAATGCTTATGAAAACGAAAAAGAAATTAATTAAGAGTAATTAATAAGAAAC 300
DB 3352 AGTGTCTAGAAATGCTTATGAAAACGAAAAAGAAATTAATTAAGAGTAATTAATAAGAAAC 3411
QY 301 ACTCATTTCTCCAGAGAGCAAGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
DB 3412 ACTCATTTCTCCAGAGAGCAAGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3471
QY 361 TTTCTAATTTCAAGAGATATAATTAATTAAGAGTAATTAATAAGAGTAATTAATAAGAGTAAT 420
DB 3472 TTTCTAATTTCAAGAGATATAATTAATTAAGAGTAATTAATAAGAGTAATTAATAAGAGTAAT 3531
QY 421 AGTGTCTGGAAGGTTCTCTGCTGT 446
DB 3532 AGTGTCTGGAAGGTTCTCTGCTGT 3557

RESULT 3
AC079775 171987 bp DNA linear PRI 29-MAY-2002
LOCUS Homo sapiens BAC clone RP11-295P2 from 2, complete sequence.
DEFINITION AC079775
ACCESSION AC079775
VERSION AC079775.6 GI:19848453
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Sulston,J.B. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2
AUTHORS Belter,E., Haakenson,W., Doebber,A. and Elliott,G.
TITLE The sequence of Homo sapiens BAC clone RP11-295P2
JOURNAL Unpublished (2001)
REFERENCE 3
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 30, 2002 this sequence version replaced gi:16924154.
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
----- Summary Statistics
```

Center project name: H\_NH0295P02

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Ooegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-261B8; the clone sequenced to the right is RP11-436K12. Actual start of this clone is at base position 1 of RP11-295P2; actual end is at base position 171987 of RP11-295P2.

#### FEATURES

repeat_region	3577..3886	/rpt_family="Alu"
repeat_region	4004..4321	/rpt_family="Alu"
repeat_region	4322..4626	/rpt_family="ERV1"
repeat_region	4629..4703	/rpt_family="Alu"
misc_feature	4669..4695	/note="match to EST BG582215 (NID:g14384950)"
repeat_region	5895..5994	/rpt_family="MIR"
repeat_region	6612..6975	/rpt_family="MaLR"
misc_feature	7608..8109	/note="match to EST BF105129 (NID:g10887655)"
repeat_region	7625..7727	/rpt_family="MIR"
repeat_region	7735..7876	/rpt_family="MER1_type"
repeat_region	8217..8547	/rpt_family="L2"
repeat_region	8984..9245	/rpt_family="Alu"
repeat_region	9251..9279	/rpt_family="(CAAA)n"
repeat_region	9283..9334	/rpt_family="Alu"
repeat_region	9335..9640	/rpt_family="Alu"
repeat_region	9641..9800	/rpt_family="Alu"
repeat_region	9802..10082	/rpt_family="Alu"
repeat_region	10088..10118	/rpt_family="(TAAA)n"
repeat_region	10385..10465	/rpt_family="MIR"
repeat_region	10667..10961	/rpt_family="Alu"
repeat_region	11074..11530	/rpt_family="MaLR"
repeat_region	11730..12034	/rpt_family="Alu"
repeat_region	12039..12322	/rpt_family="Alu"
repeat_region	12372..12408	/rpt_family="(TCCC)n"
repeat_region	12461..12651	/rpt_family="MIR"
repeat_region	12747..13036	/rpt_family="Alu"
repeat_region	13097..13413	/rpt_family="Alu"
repeat_region	13445..13730	/rpt_family="Alu"
repeat_region	13841..13968	/rpt_family="MIR"
repeat_region	13969..14260	/rpt_family="Alu"
repeat_region	14261..14368	/rpt_family="MIR"
repeat_region	14412..14720	/rpt_family="Alu"
repeat_region	15009..15312	/rpt_family="Alu"
misc_feature	15796..16451	/note="match to EST AW853926 (NID:g7949619)"
repeat_region	16612..16728	/rpt_family="MIR"
repeat_region	17083..17384	/rpt_family="Alu"
misc_feature	17168..17169	

```

/note="match to EST AW853926 (NID:g7949619)"
17563. .17713
/rpt_family="MIR"
repeat_region
17716. .18009
/rpt_family="Alu"
repeat_region
18010. .18118
/rpt_family="(CACTA)n"
misc_feature
18199. .18502
/note="match to EST BE504918 (NID:g9707326) hz32808.x1"
18441. .18855
/rpt_family="L2"

Query Match
Best Local Similarity 100.0%; Score 446; DB 9; Length 171987;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTAATGCCAGATCTCTGCTCACTGCAACCTCAGCCCTCCAGTAGTGGGATTACAGGC 60
Db 63454 GTAATGCCAGATCTCTGCTCACTGCAACCTCAGCCCTCCAGTAGTGGGATTACAGGC 63513

QY 61 ATGGCCACCACCGCGCTAAATTTGTATCTTTTAGTAGAGACGGGTTCTCCATGTT 120
Db 63514 ATGGCCACCACCGCGCTAAATTTGTATCTTTTAGTAGAGACGGGTTCTCCATGTT 63573

QY 121 GGTGAGCTGGTCTCGAATCTCAAACTCAGGTGATCCGCCCGCTCCGCTCCCAAAGT 180
Db 63574 GGTGAGCTGGTCTCGAATCTCAAACTCAGGTGATCCGCCCGCTCCGCTCCCAAAGT 63633

QY 181 GCTAGGATACAGGCTGAGCCACCGCGCTCAGCTGGGAACACCTTTCTTACATCTTC 240
Db 63634 GCTAGGATACAGGCTGAGCCACCGCGCTCAGCTGGGAACACCTTTCTTACATCTTC 63693

QY 241 AAGTGCTAGAAATGCTTATGAAACGAAAGAAATTAAGAGTAATTAAGAAAC 300
Db 63694 AAGTGCTAGAAATGCTTATGAAACGAAAGAAATTAAGAGTAATTAAGAAAC 63753

QY 301 ACTCATTTTCTCCCAAGAGAGCAAGATTTCTTTCTCTCTCTCTTTCTTTTCTTC 360
Db 63754 ACTCATTTTCTCCCAAGAGAGCAAGATTTCTTTCTCTCTCTCTTTCTTTTCTTC 63813

QY 361 TTTCTTAATTTCAAAGAGATATAATTAATGTCAGGTAAAGTCAAAAGTCTTTTAT 420
Db 63814 TTTCTTAATTTCAAAGAGATATAATTAATGTCAGGTAAAGTCAAAAGTCTTTTAT 63873

QY 421 AGTGTCTGGAAGTCTCTGCTGT 446
Db 63874 AGTGTCTGGAAGTCTCTGCTGT 63899

```

```

RESULT 4
AL356055/c
LOCUS
DEFINITION
Human DNA sequence from clone RP11-416A14 on chromosome 1, complete
sequence.
AL356055
VERSION
AL356055.13 GI:21537424
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 97288)
Griffiths, C.
Direct Submission
Submitted (31-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
On Jun 21, 2002 this sequence version replaced gi:19572362.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with

```

only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
RP11-416A14 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6.

# FEATURES

```

Source
1. .97288
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RPCI-11.2"

```

# ORIGIN

```

Query Match
Best Local Similarity 38.9%; Score 173.4; DB 9; Length 97288;
Matches 189; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 GTAATGGCAGCATCTCTGCTCACTGCAACCTCAGCCCTCCAGTAGTGGGATTACAGGC 60
Db 78382 GGAATGGTGAGATCTCGGCTCACTGCAACCTCCGCTCCCAAGTAGTGGGATTACAGGT 78323

QY 61 ATGCGCCACCAAGCCCGGCTAAATTTGTATCTTTTAGTAGAGACGGGTTCTCCATGTT 120
Db 78322 ATGTGCCCATACATCTGCTAAATTTGTATTTTGTAGTAGACAGAGGTTTCTCCATGTT 78263

QY 121 GGTGAGCTGGTCTCGAATCTCAAACTCAGGTGATCCGCCCGCTCCGCTCCCAAAGT 180
Db 78262 GGTGAGCTGGTCTCGAATCTCTGACCTCAGGTGATCCACCGCTCCCAAAGT 78203

QY 181 GCTAGGATACAGGCTGAGCCACCGCGCTCAGCC 215
Db 78202 GCTGGGATTACAGGTGTGAGCCACTGCGCAGGCC 78168

```

# RESULT 5

```

AL360268/c
LOCUS
DEFINITION
Human DNA sequence from clone RP11-379C10 on chromosome 9, complete
sequence.
AL360268
ACCESSION
AL360268.14 GI:19351891
VERSION
AL360268.14
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 110130)
Lloyd, D.
Direct Submission
Submitted (09-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
On Mar 11, 2002 this sequence version replaced gi:18650682.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations

```

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>

RP11-379C10 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-379C10 It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone RP11-379C10 is at 1 in this sequence. The true left end of clone RP11-395P17 is at 108131 in this sequence. The true right end of clone RP11-203J24 is at 303 in this sequence.

FEATURES

source 1..110130

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="9"

/clone\_lib="RP11-379C10"

/clone\_lib="RPCI-11.2"

64768

misc\_feature

/note="Tandem repeat. Forced join. Gap size estimated to be approximately 200bp by restriction digest data."

ORIGIN

Query Match 38.6%; Score 172.2; DB 9; Length 110130;

Best Local Similarity 87.3%; Pred No. 1.5e-30;

Matches 200; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

Qy 1 GTAATGCCAGATCTCTGCTCACTGCAACCTCAGCCCTCCCGAGTAGCTGGGATTACAGGC 60

Db 74369 GCAATGGCATGATCGGCTCACTGCTGCCTCAGCCCTCCCAAGTAGCTGGGATTACAGGC 74310

Qy 61 ATGGCCACACGCCCGCGCTAATTTGTATCTTTTAGTAGAGAGCGGCTTCCTCCATGTT 120

Db 74309 ATGGCCACCATCGCGCGCTAATTTGTAT-TTTTAATAGAGAGAGGTTTCTCCATGTT 74251

Qy 121 GGTGAGCTGTCTCGAACTTCAACCTCAGTGATCGCCGCGCTCGGCTCCCAAAGT 180

Db 74250 GGTGAGCTGTCTCGAACTTCTGACCTCAGGTAATCACCCACTCGGCTCCCAAAGT 74191

Qy 181 GCTAGGATTACAGCGTGTAGCCACCGCGCTCAGCGCTCGGAACACCTTTT 229

Db 74190 GCTGGATTACAAGCGTGAGCCACCGCGCGCGGGGTGCGCTTCTT 74142

RESULT 6

AL391237/c

LOCUS

DEFINITION

Human DNA sequence from clone RP13-125M24 on chromosome Xq23-25

Contains part of the gene for KIAA1058 protein, an EST, an STS and

GSSs, complete sequence.

AL391237

ACCESSION

AL391237.12 GI:12057239

HTG; KIAA1058.

KEYWORDS

source Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 58723)

AUTHORS

Wilson,S.

TITLE

Direct Submission

JOURNAL

Submitted (26-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk) requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)

COMMENT

On Jan 9, 2001 this sequence version replaced gi:10716346. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

IMPORTANT: This sequence is not the entire insert of clone RP13-125M24 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP13-125M24 is at 58723 in this sequence. The true left end of clone RP13-12804 is at 26435 in this sequence. The true right end of clone RP6-155F9 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

RP13-125M24 is from the library RPCI-13.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

FEATURES

Location/Qualifiers

source 1..58723

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="X"

/map="q23-25"

/clone="RP13-125M24"

/clone\_lib="RPCI-13.1"

58..281

repeat\_region

/note="LIMB8 repeat: matches 5645..5864 of consensus"

misc\_feature

1180..1509

/note="match: GSS: Em:AQ110399"

join(1185..1369,1461..1526,3258..3350,6262..6385,10479..10585,11078..11164,12187..12276,18286..18404,21120..21189,21279..21404,22326..22397,23320..23466,27740..27823,28648..28759,31598..31776,37587..37703,41211..41320,43143..43233,43427..43538)

/gene="b8125M24.1"

CDS

join(1185..1369,1461..1526,3258..3350,6262..6385,10479..10585,11078..11164,12187..12276,18286..18404,21120..21189,21279..21404,22326..22397,23320..23466,27740..27823,28648..28759,31598..31776,37587..37703,41211..41320,43143..43233,43427..43538)

/gene="b8125M24.1"

/note="match: cDNAs: Em:AB028981

/note="match: ESTs: Em:BE910847

match: proteins: Tr:Q9UPU4"

/codon\_start=3

repeat_region	/evidence=not experimental	/note="AluSg repeat: matches 2. .306 of consensus"
	/product="bB125M24.1 (KIAA1058 protein)"	19414. .19496
repeat_region	/protein_id="CAC36036.1"	/note="MER3 repeat: matches 119. .204 of consensus"
	/db_xref="GI:13359023"	19787. .19971
repeat_region	/db_xref="TrEMBL:Q9BMX9"	/note="MER5A repeat: matches 1. .189 of consensus"
	/translations="CITSSVPLKPFKKNQONITVEEFVPMTKYCPFTYIKNHL	20049. .20150
repeat_region	YVYPLQKQDSQTKFAKARNIACVFRDSDSDAGALKCIYGKPAGSVFTTNAYAVV	/note="51 copies 2 mer tt 60% conserved"
	SHNQPFVEDEIKIELPIJLHOKHLLFTFHVSCIEINTKGTTKQDVTVPVGEAV	20153. .20462
repeat_region	VPLLKGRITTFPQQLPVSANLPGLVNLNDQNVDIKWVDGAKPLKIKSHLESTIY	/note="AluSg repeat: matches 1. .309 of consensus"
	QDLVHAFHHKCOLQSGSKVEPGLIKYLKHAWEIQVMIQFILPVLMOQLFRVLT	20467. .20731
repeat_region	NMTHEDVPINCTMWLLHIVSKCHEGLDSYLSFKYFRPEKPSAPQAQIHTELA	/note="AluY repeat: matches 27. .302 of consensus"
	TTMIALIKQADFSLINKLKYSWFFFEITIAKSMATVLEENKIKLPRGORFFETVHH	21608. .21797
repeat_region	SPKDPKVLAVKPEFLOTICNHEHYIPLNLPMAFPKLPQVQDSNLEYSLSDEYKX	/note="MIR repeat: matches 19. .262 of consensus"
	HFLVGLLRSTSLALQDNYEIRTAISVINKLLIKHAFDTRYOHKQQAQKIAQLYLPF	21931. .22096
repeat_region	GLULENIQRLAGRDITLSCAAMPNSLGNLFRKMBELREKIQEVPSQKEQDFQIRA	/note="MER97b repeat: matches 2. .177 of consensus"
	TLVKIHDFKILPLKMAVENETQYSLYLCEDSKTRQSSVSVQYNRLDQYEIRS	23573. .23747
repeat_region	LMCYLYIVKMISE"	/note="L1M1 repeat: matches 5581. .5758 of consensus"
	3602. .3700	23805. .24103
repeat_region	/note="MIR repeat: matches 46. .146 of consensus"	/note="AluSx repeat: matches 1. .299 of consensus"
	4610. .4693	24184. .24312
repeat_region	/note="37 copies 2 mer tt 64% conserved"	/note="L1M1 repeat: matches 5457. .5582 of consensus"
	4691. .5493	24787. .25049
repeat_region	/note="L1PA2 repeat: matches 5342. .6144 of consensus"	/note="AluSx repeat: matches 44. .312 of consensus"
	6498. .6849	26068. .26311
repeat_region	/note="THE1A repeat: matches 1. .354 of consensus"	/note="L2 repeat: matches 2470. .2710 of consensus"
	7361. .7390	26460. .27076
repeat_region	/note="L2 repeat: matches 2681. .2710 of consensus"	/gene="BB125M24.1"
	7391. .7617	/note="match: GSS: Em:AQ532152"
repeat_region	/note="MER46A repeat: matches 1. .236 of consensus"	26741. .26868
	7618. .8526	/note="FRAM repeat: matches 1. .145 of consensus"
repeat_region	/note="L2 repeat: matches 1710. .2681 of consensus"	26937. .27193
	8904. .9133	/note="L2 repeat: matches 2008. .2286 of consensus"
repeat_region	/note="L1MB8 repeat: matches 5940. .6169 of consensus"	27439. .27496
	11296. .11457	/note="29 copies 2 mer tg 75% conserved"
repeat_region	/note="FRAM repeat: matches 2. .161 of consensus"	27980. .28521
	11516. .11698	/note="MER88B repeat: matches 8. .569 of consensus"
repeat_region	/note="MIR repeat: matches 23. .197 of consensus"	29573. .29608
	12983. .13173	/note="18 copies 2 mer tt 88% conserved"
repeat_region	/note="MIR repeat: matches 3. .215 of consensus"	30097. .30187
	13174. .13485	/note="L2 repeat: matches 2642. .2731 of consensus"
repeat_region	/note="AluSg repeat: matches 1. .308 of consensus"	30258. .30552
	13486. .13517	/note="AluDb repeat: matches 1. .295 of consensus"
repeat_region	/note="MIR repeat: matches 215. .243 of consensus"	30649. .30950
	13511. .14151	/note="AluSx repeat: matches 3. .302 of consensus"
repeat_region	/note="L2 repeat: matches 2055. .2748 of consensus"	31914. .32206
	14190. .14492	/note="AluSx repeat: matches 1. .292 of consensus"
repeat_region	/note="AluSx repeat: matches 1. .305 of consensus"	32240. .32546
	14821. .15137	/note="AluDb repeat: matches 1. .300 of consensus"
repeat_region	/note="AluSx repeat: matches 1. .305 of consensus"	
	15162. .15241	
repeat_region	/note="MIR repeat: matches 58. .140 of consensus"	
	15318. .15477	
repeat_region	/note="L2 repeat: matches 2570. .2746 of consensus"	
	15595. .15888	
repeat_region	/note="AluSx repeat: matches 1. .294 of consensus"	
	15911. .16504	
repeat_region	/note="L2 repeat: matches 2080. .2710 of consensus"	
	16880. .16937	
repeat_region	/note="L2 repeat: matches 2326. .2384 of consensus"	
	17031. .17241	
repeat_region	/note="MER20 repeat: matches 1. .218 of consensus"	
	17467. .17566	
repeat_region	/note="L2 repeat: matches 2594. .2698 of consensus"	
	17560. .17597	
repeat_region	/note="19 copies 2 mer tt 89% conserved"	
	17570. .17597	
repeat_region	/note="14 copies 2 mer tt 100% conserved"	
	18437. .18993	
repeat_region	/note="L1 repeat: matches 2472. .3053 of consensus"	
	19010. .19129	
repeat_region	/note="MER3 repeat: matches 9. .166 of consensus"	
	19130. .19413	

Query Match

Best Local Similarity 38.5%; Score 171.6; DB 9; Length 58723;

Mismatches 0; Gaps 0;

QY	1	GTAAATGGCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCCAGCTAGCTGGGATTACAGGC	60
Db	58674	GCAGTGGCGCGATCTTGCTCACTGCACCTCCGCTCCTGAGTAGCTGGGATTACAGGC	58615
QY	61	ATGCGCCACACGCCCGGCTAAATTTGTATCTTTTAGTAGACGGCGTTCTCCATGTT	120
Db	58614	ACCCGCCACACGCCCGGCTAAATTTGTATCTTTTAGTAGAGATGGGTTTCCCATGTT	58555
QY	121	GCTCAGGCTGGTCTCGAACTTCAACCTCAGGTGATCCGCCCGCTCCGCTCCCAAACT	180
Db	58554	GCTCAGGCTGGTGTCAAACTCTGACCTCAGGTGATCTGCCCGCTCCGCTCCCAAACT	58495
QY	181	GCTAGGATTACAGGCGTGAGCCACCGCTCAGCCTGGGAAACACCTTTTCTTACATCTTC	240
Db	58494	GCTGGGATTATAGGCGTGAACCACTGCGCGGCTCATGATTACTTAATTTCTTTTATT	58435
QY	241	AGTCTCTAGAATGCTTATGAAAAACGAAAAAGAAATTAATA	282
Db	58434	TATTTTCTCTCCCTATCCCACTAAGACTGAACGCTTCA	58393

## RESULT 7

AC021763/c  
 LOCUS AC021763 158405 bp DNA linear PRI 23-AUG-2002  
 DEFINITION Homo sapiens chromosome 18, clone RP11-56021, complete sequence.  
 ACCESSION AC021763  
 VERSION AC021763.10 GI:22450664  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 158405)

TITLE Birren,B., Nusbaum,C. and Lander,E.  
 JOURNAL Homo sapiens chromosome 18, clone RP11-56021  
 AUTHORS Unpublished

REFERENCE  
 AUTHORS 2 (bases 1 to 158405)  
 TITLE Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
 Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenesstor,J.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
 McPheters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,  
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.  
 JOURNAL Direct Submission

REFERENCE  
 AUTHORS Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 158405)

TITLE Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 JOURNAL Direct Submission

REFERENCE  
 AUTHORS Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 158405)

TITLE Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 JOURNAL Direct Submission

REFERENCE  
 AUTHORS Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 158405)

TITLE Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 JOURNAL Direct Submission

Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 JOURNAL Direct Submission

COMMENT Submitted (23-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 23, 2002 this sequence version replaced gi:22123424.  
 All repeats were identified using RepeatMasker:  
 Smit, A.P.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L2274  
 Center clone name: 56\_O\_21  
 ----- Location/Qualifiers  
 source 1. 158405  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="18"  
 /map="18"  
 /clone="RP11-56021"  
 /clone\_lib="RPCI-11 Human Male BAC"  
 complement(13. 217)  
 /rpt\_family="AluSx"  
 993. 1363  
 /rpt\_family="AluSx"  
 1733. 1920  
 /rpt\_family="MIR"  
 complement(4542. 4743)  
 /rpt\_family="AluSg"  
 complement(5403. 5566)  
 /rpt\_family="AluJb"  
 complement(5567. 5861)  
 /rpt\_family="AluSg"  
 complement(5862. 5996)  
 /rpt\_family="AluJb"  
 6220. 6364  
 /rpt\_family="AluJb"  
 complement(6365. 6671)  
 /rpt\_family="AluSx"  
 6672. 6853  
 /rpt\_family="AluJb"  
 complement(6875. 6995)  
 /rpt\_family="L2"  
 complement(7249. 7554)  
 /rpt\_family="AluSx"  
 7584. 7858  
 /rpt\_family="AluJb"  
 complement(7899. 8188)  
 /rpt\_family="AluSx"  
 complement(8572. 8870)  
 /rpt\_family="AluSg"  
 complement(9304. 9370)  
 /rpt\_family="L2"  
 9371. 9677  
 /rpt\_family="AluSg"  
 complement(9678. 9715)  
 /rpt\_family="L2"  
 9714. 9753  
 /rpt\_family="MIR3"  
 9926. 10206  
 /rpt\_family="AluSg"  
 10381. 10677  
 /rpt\_family="AluY"  
 10717. 10740





175082 177828 contig of 2747 bp in length  
 177929 180660 contig of 2732 bp in length  
 180761 183132 contig of 2432 bp in length  
 183293 185028 contig of 1736 bp in length  
 185129 185875 contig of 747 bp in length  
 185976 187200 contig of 1225 bp in length.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 21 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 33581: contig of 33581 bp in length  
 33681: gap of 100 bp  
 33682 38093: contig of 24412 bp in length  
 38094 58193: gap of 100 bp  
 58194 77625: contig of 19432 bp in length  
 77626 97031: gap of 100 bp  
 97032 97131: contig of 19306 bp in length  
 97132 97132: gap of 100 bp  
 111990: contig of 14859 bp in length  
 111991 12090: gap of 100 bp  
 12091 125465: contig of 13375 bp in length  
 125466 125465: gap of 100 bp  
 125466 135454: contig of 9889 bp in length  
 135455 135554: gap of 100 bp  
 135555 144628: contig of 9074 bp in length  
 144629 144728: gap of 100 bp  
 144729 151062: contig of 6334 bp in length  
 151063 151163: gap of 100 bp  
 151163 15511: contig of 4349 bp in length  
 15512 155611: gap of 100 bp  
 155612 160186: contig of 4575 bp in length  
 160187 160286: gap of 100 bp  
 160287 164088: contig of 3802 bp in length  
 164089 164188: gap of 100 bp  
 164189 167739: contig of 3551 bp in length  
 167740 167839: gap of 100 bp  
 167840 171188: contig of 3349 bp in length  
 171189 171288: gap of 100 bp  
 171289 174981: contig of 3693 bp in length  
 174982 175081: gap of 100 bp  
 175082 177828: contig of 2747 bp in length  
 177829 177928: gap of 100 bp  
 177929 180660: contig of 2732 bp in length  
 180661 180760: gap of 100 bp  
 180761 183192: contig of 2432 bp in length  
 183193 183292: gap of 100 bp  
 183293 185028: contig of 1736 bp in length  
 185029 185128: gap of 100 bp  
 185129 185875: contig of 747 bp in length  
 185876 185975: gap of 100 bp  
 185976 187200: contig of 1225 bp in length.

Location/Qualifiers  
 1. .187200  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="18"  
 /map="18q12"  
 /clone="RP11-850H3"

1. .33581  
 /note="assembly\_fragment"  
 33682. .58093  
 /note="assembly\_fragment"  
 58194. .77625  
 /note="assembly\_fragment"  
 77626. .97031  
 /note="assembly\_fragment clone\_end:SP6 vector\_side:right"  
 97132. .111990  
 /note="assembly\_fragment"  
 111991. .12090  
 /note="assembly\_fragment"

misc\_feature  
 misc\_feature  
 misc\_feature  
 misc\_feature  
 misc\_feature

misc\_feature 112091. .125465  
 /note="assembly\_fragment"  
 misc\_feature 125466. .135454  
 /note="assembly\_fragment"  
 misc\_feature 135455. .144628  
 /note="assembly\_fragment"  
 misc\_feature 144729. .151062  
 /note="assembly\_fragment"  
 misc\_feature 151163. .15511  
 /note="assembly\_fragment"  
 misc\_feature 155612. .160186  
 /note="assembly\_fragment"  
 misc\_feature 160287. .164088  
 /note="assembly\_fragment"  
 misc\_feature 164189. .167739  
 /note="assembly\_fragment"  
 misc\_feature 167840. .171188  
 /note="assembly\_fragment"  
 misc\_feature 171289. .174981  
 /note="assembly\_fragment"  
 misc\_feature 175082. .177828  
 /note="assembly\_fragment"  
 misc\_feature 177929. .180660  
 /note="assembly\_fragment"  
 misc\_feature 180761. .183192  
 /note="assembly\_fragment"  
 misc\_feature 183293. .185028  
 /note="assembly\_fragment"  
 misc\_feature 185129. .185875  
 /note="assembly\_fragment clone\_end:T7 vector\_side:right"  
 185976. .187200  
 /note="assembly\_fragment"

ORIGIN  
 Query Match 38.4%; Score 171.4; DB 2; Length 187200;  
 Best Local Similarity 82.7%; Pred. No. 2.6e-30;  
 Matches 196; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1 GTAATGGCAGATCTCTGCTCACTGCAACCTCAGCCCTCCCGAGTAGCTGGGATTACAGGC 60  
 Db 169862 GCAGTGGCGTGTCTAGCTCAGCGCACTCAGACTCCCGAGTAGCTGGGATTACAGGC 169921

QY 61 ATCGCCACACACCCCGGCTAATTTGTATCTTTAGTAGAGACGGCGTTCTCCATGTT 120  
 Db 169922 ATCGCCACACACCGCTGGCTAATTTGTGTCTTTACTAGAGACGGGTTCTCCATGTT 169981

QY 121 GCTCAGGCTGGTCTCGAATCTCAACCTCAGGTATCCGCCCGCTCCGCTCCCAAGT 180  
 Db 169982 GGTGAGGCTGGTCTTGAATCCCGGACTTCATGTGATCCACCCGCTCCCAAGT 170041

QY 181 GCTAGGATTACAGCGGTGAGCCACCGGCTCAGCTGGGAACACCTTTCTTACATC 237  
 Db 170042 GCTGGGATTACAGCGGTGAGCCACCGGCTCCGCTCCGCTCCCAAGT 170098

RESULT 9  
 AC064841  
 LOCUS AC064841 69829 bp DNA linear PRI 21-FEB-2002  
 DEFINITION Homo sapiens BAC clone CTD-2015A10 from 2, complete sequence.  
 ACCESSION AC064841  
 VERSION AC064841.3 GI:18464243  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 69829)  
 AUTHORS Sulston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 69829)

# AUTHORS TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

Abbott, S. and Cotton, M.  
The sequence of Homo sapiens BAC clone CTD-2015A10  
Unpublished (2001)  
3 (bases 1 to 69829)  
Waterson, R.H.  
Direct Submission  
Submitted (23-APR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 69829)  
Waterson, R.H.  
Direct Submission  
Submitted (01-FEB-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 69829)

## REFERENCE AUTHORS TITLE JOURNAL

Waterson, R.  
Direct Submission  
Submitted (21-FEB-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Feb 1, 2002 this sequence version replaced gi:18151025.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
----- Summary Statistics  
-----

## COMMENT

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

Clone CTD-2015A10 is from a release of the human BAC library CTD.  
The library contains cloned DNA from human sperm. See: Shizuya et  
al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al.,  
Genomics 34:213-8 (1996). The clone is available from Research  
Genetics, Inc. (<http://www.resgen.com>).  
VECTOR: pBelobAC11  
Selection: chloramphenicol

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-12J20, 2000 bp overlap; the  
clone sequenced to the right is RP11-765C7, 2000 bp overlap.  
Actual start of this clone is at base position 9553 of RP11-12J20;  
actual end is at base position 26527 of RP11-765C7.

## FEATURES

source

Location/Qualifiers  
1. .69829  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="CTD-2015A10"  
/clone\_lib="CTD"  
2. .160

repeat\_region

repeat\_region  
440. .612  
/rpt\_family="Alu"  
repeat\_region  
628. .1330  
/rpt\_family="L1"  
repeat\_region  
1774. .2250  
/rpt\_family="ERV1"  
repeat\_region  
3970. .4015  
/rpt\_family="(T)n"  
repeat\_region  
3986. .4296  
/rpt\_family="Alu"  
repeat\_region  
5165. .5191  
/rpt\_family="(TCTG)n"  
repeat\_region  
5328. .5362  
/rpt\_family="AT\_rich"  
repeat\_region  
7399. .7423  
/rpt\_family="AT\_rich"  
repeat\_region  
7961. .8530  
/rpt\_family="ERV1"  
repeat\_region  
8617. .8637  
/rpt\_family="AT\_rich"  
repeat\_region  
8961. .9026  
/rpt\_family="ERV1"  
repeat\_region  
9254. .9274  
/rpt\_family="AT\_rich"  
repeat\_region  
9560. .9915  
/rpt\_family="MER2\_type"  
repeat\_region  
9916. .10207  
/rpt\_family="Alu"  
repeat\_region  
9918. .9941  
/rpt\_family="(TTTTG)n"  
repeat\_region  
10208. .10306  
/rpt\_family="MER2\_type"  
repeat\_region  
10327. .10364  
/rpt\_family="AT\_rich"  
repeat\_region  
10678. .10736  
/rpt\_family="L2"  
repeat\_region  
11041. .11623  
/rpt\_family="ERV1"  
repeat\_region  
11836. .12057  
/rpt\_family="MER2\_type"  
repeat\_region  
12058. .12394  
/rpt\_family="MER1\_type"  
repeat\_region  
12395. .12515  
/rpt\_family="MER2\_type"  
repeat\_region  
12573. .12868  
/rpt\_family="Alu"  
repeat\_region  
13224. .13377  
/rpt\_family="MaLR"  
repeat\_region  
13403. .13464  
/rpt\_family="L2"  
repeat\_region  
13605. .13813  
/rpt\_family="L1"  
repeat\_region  
14097. .14148  
/rpt\_family="L2"  
repeat\_region  
14149. .14413  
/rpt\_family="Alu"  
repeat\_region  
14414. .14696  
/rpt\_family="L2"  
repeat\_region  
15133. .15180  
/rpt\_family="AT\_rich"  
repeat\_region  
15310. .15491  
/rpt\_family="L1"  
repeat\_region  
15492. .15616  
/rpt\_family="Alu"  
repeat\_region  
15617. .15715  
/rpt\_family="Alu"  
repeat\_region  
15720. .16010  
/rpt\_family="Alu"  
repeat\_region  
16038. .16753  
/rpt\_family="L1"  
repeat\_region  
18405. .18924  
/rpt\_family="L1"



```

* 16658 16757: gap of 100 bp
* 16758 26995: contig of 10238 bp in length
* 26996 27095: gap of 100 bp
* 27096 42276: contig of 15181 bp in length
* 42277 60953: contig of 18577 bp in length
* 60954 61054: gap of 100 bp
* 61054 83707: contig of 22654 bp in length
* 83708 93807: gap of 100 bp
* 93808 116924: contig of 33117 bp in length
* 116925 117025: gap of 100 bp
* 117025 178670: contig of 61646 bp in length.
FEATURES
    source
        1. .178670
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="RP11-45118"
            /clone_lib="RPC1-11 Human Male BAC"
        1. .1183
            /note="assembly_fragment"
        1284. .3307
            /note="assembly_fragment"
        3408. .7626
            /note="assembly_fragment"
        7727. .16657
            /note="assembly_fragment"
        16758. .26995
            /note="assembly_fragment"
        27096. .42276
            /note="assembly_fragment"
        42377. .60953
            /note="assembly_fragment"
        clone_end:T7
        vector_side:left"
        61054. .83707
            /note="assembly_fragment"
        83808. .116924
            /note="assembly_fragment"
        clone_end:SP6
        vector_side:left"
        117025. .178670
            /note="assembly_fragment"
    ORIGIN
        Query Match      38.3%;   Score 171;   DB 2;   Length 178670;
        Best Local Similarity 66.9%;   Pred. No. 3.2e-30;
        Matches 243;   Conservative 0;   Mismatches 120;   Indels 0;   Gaps 0;

QY      1  GTAATGCCAGCATCTGTGCTCACTGCAACCTCAGCCTCCCGCTAGCTAGCTGGATTACAGGC 60
DB      50841  GCATGGCAGATCTTGGCTCACTGTATCTCTGCTCTGAGTAGCTGGATTACAGGC 50782

QY      61  ATGGCCACACCGCCGGCTAAATTTGTATCTTTTAGTAGAGAGCGGTTCTCCATGTT 120
DB      50781  ATGCACCCACACCTGGCTAAATTTGTATTTTAGTAGAGACAGGGTTCCACCATGTT 50722

QY      121  GGTGAGCTGGTCTCGAATCTCAAACTCAGGTGATCCGCGCTCGGCTCCCAAGT 180
DB      50721  GGCAGGCTGGTCTTGAACCTCCTGAGCTCAGGTGATCTGCGCGCTCGGCTCCCAAGT 50662

QY      181  GCTAGGATTACAGCGCTGAGCCACCGCTCAGCTCAGCGGGAACACCTTTCTTACATCTTC 240
DB      50661  GCTGGGATTACAGCATGAGCCACCGTGCCTGCTGACTCTCTGTTATTTTGTGTGT 50602

QY      241  AAGTGCTAGAAATGCTTATGAAACGAAAGAAATATTAAGAGTAATATTAAGAAAC 300
DB      50601  CAGTTTAAATGCTGTGTTTTAGGAGGCGCTTTCTGCGCATCTATTAAATGTCATT 50542

QY      301  ACTCATTTTCTCCAGAGACCAAGATTCTCTTTCTCTCTCTCTTTTCTTTTTC 360
DB      50541  TCCCTTCAATTTCTGAAGCTAGTCTTGTGTTTTTACAGAACTCTTATTTTCCCTTATT 50482

```

```

QY      361  TTT 363
DB      50481  TTT 50479

RESULT 11
AC017079/c
LOCUS   AC017079          176075 bp      DNA      linear      PRI 07-NOV-2001
DEFINITION Homo sapiens BAC clone RP11-462M9 from 2, complete sequence.
ACCESSION AC017079
VERSION   AC017079.5   GI:13625506
KEYWORDS HTG.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 176075)
            Sulston,J.E. and Waterston,R.
            Toward a complete human genome sequence
            Genome Res. 8 (11), 1097-1108 (1998)
            99083792
            PUBMED 9847074
            2 (bases 1 to 176075)
            Belter,E., Maupin,R., Hawkins,M. and Le,T.
            The sequence of Homo sapiens BAC clone RP11-462M9
            Unpublished
            3 (bases 1 to 176075)
            Waterston,R.H.
            Direct Submission
            Submitted (09-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            4 (bases 1 to 176075)
            Waterston,R.H.
            Direct Submission
            Submitted (15-APR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            5 (bases 1 to 176075)
            Waterston,R.
            Direct Submission
            Submitted (09-AUG-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Apr 15, 2001 this sequence version replaced gi:13431206.
            ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            ----- Summary Statistics
            -----
            Center project name: H_NH0462M09
            -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.

```

McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPci-11 human BAC library was made from the blood of one male donor, as described by Ooegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-480M9. Actual start of this clone is at base position 1 of RP11-462M9; actual end is at base position 176075 of RP11-462M9.

The sequence H.NH0462M09 from base positions 164804 to 164876 and 165220 to 165298 are represented by sequence derived from PCR on BAC DNA.

#### FEATURES

source	Location/Qualifiers
	1. .176075
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-462M9"
	/clone_lib="RPCI-11"
	6. .262
repeat_region	/rpt_family="L1"
repeat_region	577. .624
repeat_region	/rpt_family="(CA)n"
repeat_region	1071. .1100
repeat_region	/rpt_family="AT-rich"
misc_feature	1116. .1469
misc_feature	/note="similar to EST BF329553 (NID:g11300301)"
misc_feature	2003. .2101
misc_feature	/note="similar to EST AU139468 (NID:g11000989)"
misc_feature	2003. .2101
misc_feature	/note="similar to EST AX085613 (NID:g13275620)"
misc_feature	2003. .2101
misc_feature	/note="similar to EST BE569937 (NID:g9813657)"
misc_feature	2906. .2929
misc_feature	/note="similar to EST AU139468 (NID:g11000989)"
misc_feature	2906. .2929
misc_feature	/note="similar to EST AX085613 (NID:g13275620)"
repeat_region	3070. .3382
repeat_region	/rpt_family="Alu"
misc_feature	3521. .3864
misc_feature	/note="similar to EST BF901906 (NID:g12293365)"
misc_feature	3766. .3864
misc_feature	/note="similar to EST AU139468 (NID:g11000989)"
misc_feature	3766. .3864
misc_feature	/note="similar to EST AX085613 (NID:g13275620)"
misc_feature	3766. .3864
misc_feature	/note="similar to EST BE569937 (NID:g9813657)"
misc_feature	3792. .3864
misc_feature	/note="similar to EST BF159650 (NID:g11039755)"
misc_feature	3797. .3864
misc_feature	/note="similar to EST AA589501 (NID:g2402881) v148c05.s1"
misc_feature	3810. .3864
misc_feature	/note="similar to EST BG247212 (NID:g12757027)"
repeat_region	4727. .5223
repeat_region	/rpt_family="L1"
repeat_region	5495. .5607
repeat_region	/rpt_family="L2"
repeat_region	5608. .5911
repeat_region	/rpt_family="Alu"
repeat_region	5912. .5947
repeat_region	/rpt_family="L2"

misc_feature	6028. .6235	/note="similar to EST BE569937 (NID:g9813657)"
misc_feature	6028. .6234	/note="similar to EST AA589501 (NID:g2402881) v148c05.s1"
misc_feature	6028. .6234	/note="similar to EST AU139468 (NID:g11000989)"
misc_feature	6028. .6234	/note="similar to EST AX085613 (NID:g13275620)"
misc_feature	6028. .6234	/note="similar to EST BF159650 (NID:g11039755)"
misc_feature	6028. .6234	/note="similar to EST BG247212 (NID:g12757027)"
misc_feature	6028. .6121	/note="similar to EST BF901906 (NID:g12293365)"
misc_feature	6143. .6234	/note="similar to EST BF370277 (NID:g11332302)"
misc_feature	6188. .6234	/note="similar to EST BG173033 (NID:g12679840)"
misc_feature	6973. .7105	/note="similar to EST BF880772 (NID:g12270898)"
misc_feature	6974. .7105	/note="similar to EST AA589501 (NID:g2402881) v148c05.s1"
misc_feature	6974. .7105	/note="similar to EST AX085613 (NID:g13275620)"
misc_feature	6974. .7105	/note="similar to EST BF159650 (NID:g11039755)"
misc_feature	6974. .7105	/note="similar to EST BF370277 (NID:g11332302)"
misc_feature	6974. .7105	/note="similar to EST BG173033 (NID:g12679840)"
misc_feature	6974. .7105	/note="similar to EST BG247212 (NID:g12757027)"
misc_feature	6974. .7065	/note="similar to EST AU139468 (NID:g11000989)"
misc_feature	7845. .7782	/note="similar to EST AA333308 (NID:g1985551)"
misc_feature	7852. .7782	/note="similar to EST BF370277 (NID:g11332302)"
misc_feature	7852. .7782	/note="similar to EST AX085613 (NID:g13275620)"
misc_feature	7852. .7782	/note="similar to EST BF159650 (NID:g11039755)"
misc_feature	7852. .7782	/note="similar to EST BF880772 (NID:g12270898)"
misc_feature	7852. .7782	/note="similar to EST BG173033 (NID:g12679840)"
misc_feature	7852. .7782	/note="similar to EST BG247212 (NID:g12757027)"
repeat_region	8068. .8231	/rpt_family="L1"
repeat_region	8412. .8855	/rpt_family="L1"
repeat_region	8887. .8965	/rpt_family="Alu"
misc_feature	9364. .9641	/note="similar to EST AW856912 (NID:g7952605)"
repeat_region	9948. .10240	/rpt_family="Alu"
repeat_region	10279. .10410	

Query Match 38.3%; Score 170.6; DB 9; Length 176075;  
Best Local Similarity 69.8%; Pred. No. 4e-30;  
Matches 259; Conservative 0; Mismatches 109; Indels 3; Gaps 2;

Qy	23	CTGCAACCTCAGCCTCCCGGATTACAGGCATGCGCCACCGCGGTAA	82
Db	91905	CTCCTGCTCAGCCTCCCTAGCAGCTAGGATTACAGGCATGCGCCACCGCGGTAA	91846
Qy	83	TTTTGTATCTTTTAGTAGAGACGGCGTTCTTCATGTTGGTCAAGCTGTCGAATTC	142
Db	91845	TTTTGTAT-GTTTAGTAGAGACGGGTTTCTCCATGTTGGTCAAGCTGTCGAATTC	91787

```
QY 143 AACCTCAGTGATCCGCCGCTCGGCTCCCAAGTCTAGATTACAGGCGTGAGCC 202
Db 91786 TGACCTCAGGTGATCCACCCACCTTGGCTCCCAAAAGTCTGGATTACAGGCGTGAGCC 91727

QY 203 ACCGCGCTCAGCGCTGGGAACACCTTTTCTTACATCTTCAAGTGTAGAAATGCTTATGAA 262
Db 91726 TCCGGCGCCGCCAACAATCTCTATATTTTAATTTGGTGATTTAAACCACTGATGTTCAA 91667

QY 263 AAGCAAAAAGAAATTTAAGAGTAATATAAAGAAACACTCAAT--TTCCTCCCAAGAG 320
Db 91666 AGTGAATATTCATATAGATTAGATTAAATATTACCATATATTTTACTATTTTCTATTAG 91607

QY 321 AGCAAGATTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 380
Db 91606 TTGCCCTCGTCTTCATGCGCTATTTTGGCTTCCACTCTTTTCTGCGCTTTTGAGGTTT 91547

QY 381 TAATTAATTTG 391
Db 91546 AATTGAGATTG 91536

RESULT 12
AX871940
LOCUS
DEFINITION Sequence 6845 from Patent EP1074617.
ACCESSION AX871940
VERSION AX871940.1 GI:40026767
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primers for synthesizing full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 6845 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source 1..568
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 38.1%; Score 170; DB 6; Length 568;
Best Local Similarity 81.4%; Pred. No. 2.5e-30;
Matches 197; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 23 CTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCCGCTAA 82
Db 91 CTCCGCGCTCAGCTTCCCAAGTAGCTGGGATTACAGGCATGCGCCACCGCTGCTGCTAA 150

QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTCTCCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142
Db 151 TTTTGGTATTTTTCAGTAGAGACAGGGTTTCTCCATGTTAGTCAGGCTGGTCTCGAACTCC 210

QY 143 AAACCTCAGGTGATCCGCCGCTCGGCTCCCAAGTCTAGGATTACAGGCGTGAGCC 202
Db 211 CGACCTCAGATGATCCGCCCTCCCAAGTCTCGGCTCCCAAGTCTGGGATTACAGGCGTGAGCC 270

QY 203 ACCGCGCTCAGCTGGGAACACCTTTTCTTACATCTTCAAGTGTAGAAATGCTTATGAA 262
Db 271 ACCAGCCCGAGCTTAAAGAAATCTTTAAATAATTTTCTGGTGCTCTACATGTTACAGAGA 330

QY 263 AA 264
Db 331 AA 332

RESULT 13
BD152002
LOCUS
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD152002
VERSION BD152002.1 GI:27857760
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primers for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 6845 09-JUL-2002;
HELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/6845
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10, C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1..568
/organism="Homo sapiens (human)".
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source 1..568
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 38.1%; Score 170; DB 6; Length 568;
Best Local Similarity 81.4%; Pred. No. 2.5e-30;
Matches 197; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 23 CTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCCGCTAA 82
Db 91 CTCCGCGCTCAGCTTCCCAAGTAGCTGGGATTACAGGCATGCGCCACCGCTGCTGCTAA 150

QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTCTCCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142
Db 151 TTTTGGTATTTTTCAGTAGAGACAGGGTTTCTCCATGTTAGTCAGGCTGGTCTCGAACTCC 210

QY 143 AAACCTCAGGTGATCCGCCGCTCGGCTCCCAAGTCTAGGATTACAGGCGTGAGCC 202
Db 211 CGACCTCAGATGATCCGCCCTCCCAAGTCTCGGCTCCCAAGTCTGGGATTACAGGCGTGAGCC 270

QY 203 ACCGCGCTCAGCTGGGAACACCTTTTCTTACATCTTCAAGTGTAGAAATGCTTATGAA 262
Db 271 ACCAGCCCGAGCTTAAAGAAATCTTTAAATAATTTTCTGGTGCTCTACATGTTACAGAGA 330

QY 263 AA 264
Db 331 AA 332

RESULT 14
AC144522/c
LOCUS
DEFINITION Homo sapiens 12 BAC RP11-481J8 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.
ACCESSION AC144522
VERSION AC144522.12 GI:45592967
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```



```
AC026856
LOCUS       AC026856               165593 bp    DNA    linear    HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 12 clone RP11-237C13, WORKING DRAFT
ACCESSION   AC026856
VERSION     AC026856.2   GI:7577612
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 165593)
            Waterston,R.H.
            The sequence of Homo sapiens clone
            Unpublished
  REFERENCE 2 (bases 1 to 165593)
            Waterston,R.H.
            Direct Submission
            Submitted (24-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Apr 16, 2000 this sequence version replaced gi:7321947.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0237C13
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Assembly: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158514 bases at least Q40
Consensus quality: 160632 bases at least Q30
Consensus quality: 161787 bases at least Q20
Insert size: 167000; agarose-fp
Quality coverage: 5.73 in Q20 bases; agarose-fp
Quality coverage: 5.88 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1609: contig of 1609 bp in length
* 1610 1709: gap of unknown length
* 1710 3292: contig of 1583 bp in length
* 3293 3392: gap of unknown length
* 3393 5714: contig of 2322 bp in length
* 5715 5814: gap of unknown length
* 5815 8650: contig of 2836 bp in length
* 8651 8751: contig of 2713 bp in length
* 8751 11463: contig of 3605 bp in length
* 11464 11563: gap of unknown length
* 11564 15168: contig of 3605 bp in length
* 15169 15269: contig of 5641 bp in length
* 15269 20910: gap of unknown length
* 20910 26741: contig of 5732 bp in length
* 26741 26842: gap of unknown length
* 26842 32213: contig of 5372 bp in length
* 32214 32314: gap of unknown length
* 32314 39059: contig of 6746 bp in length
* 39060 39159: gap of unknown length
* 39160 47572: contig of 8413 bp in length
* 47573 47672: gap of unknown length
```

```
AC026856
LOCUS       AC026856               165593 bp    DNA    linear    HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 12 clone RP11-237C13, WORKING DRAFT
ACCESSION   AC026856
VERSION     AC026856.2   GI:7577612
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 165593)
            Waterston,R.H.
            The sequence of Homo sapiens clone
            Unpublished
  REFERENCE 2 (bases 1 to 165593)
            Waterston,R.H.
            Direct Submission
            Submitted (24-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Apr 16, 2000 this sequence version replaced gi:7321947.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0237C13
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Assembly: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158514 bases at least Q40
Consensus quality: 160632 bases at least Q30
Consensus quality: 161787 bases at least Q20
Insert size: 167000; agarose-fp
Quality coverage: 5.73 in Q20 bases; agarose-fp
Quality coverage: 5.88 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1609: contig of 1609 bp in length
* 1610 1709: gap of unknown length
* 1710 3292: contig of 1583 bp in length
* 3293 3392: gap of unknown length
* 3393 5714: contig of 2322 bp in length
* 5715 5814: gap of unknown length
* 5815 8650: contig of 2836 bp in length
* 8651 8751: contig of 2713 bp in length
* 8751 11463: contig of 3605 bp in length
* 11464 11563: gap of unknown length
* 11564 15168: contig of 3605 bp in length
* 15169 15269: contig of 5641 bp in length
* 15269 20910: gap of unknown length
* 20910 26741: contig of 5732 bp in length
* 26741 26842: gap of unknown length
* 26842 32213: contig of 5372 bp in length
* 32214 32314: gap of unknown length
* 32314 39059: contig of 6746 bp in length
* 39060 39159: gap of unknown length
* 39160 47572: contig of 8413 bp in length
* 47573 47672: gap of unknown length
```



\* 47673 57070: contig of 9398 bp in length  
 \* 57071 57170: gap of unknown length  
 \* 57171 68217: contig of 11047 bp in length  
 \* 68218 68317: gap of unknown length  
 \* 68318 79880: contig of 11563 bp in length  
 \* 79881 79980: gap of unknown length  
 \* 79981 95382: contig of 15402 bp in length  
 \* 95383 95482: gap of unknown length  
 \* 95483 115764: contig of 20282 bp in length  
 \* 115765 115865: gap of unknown length  
 \* 115866 141797: contig of 25933 bp in length  
 \* 141798 141897: gap of unknown length  
 \* 141898 165593: contig of 23696 bp in length.

## FEATURES

## source

1. .165593  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="12"  
 /clone="RP11-237C13"  
 misc\_feature 1. .1609  
 /note="assembly\_name:Contig3"  
 misc\_feature 1710. .3292  
 /note="assembly\_name:Contig4"  
 misc\_feature 3393. .5714  
 /note="assembly\_name:Contig5"  
 misc\_feature 5815. .8650  
 /note="assembly\_name:Contig6"  
 misc\_feature 8751. .11463  
 /note="assembly\_name:Contig7"  
 misc\_feature 11564. .15168  
 /note="assembly\_name:Contig8"  
 misc\_feature 15269. .20909  
 /note="assembly\_name:Contig9"  
 misc\_feature 21010. .26741  
 /note="assembly\_name:Contig10"  
 misc\_feature 26842. .32213  
 /note="assembly\_name:Contig11"  
 misc\_feature 32314. .39059  
 /note="assembly\_name:Contig12"  
 misc\_feature 39160. .47572  
 /note="assembly\_name:Contig13"  
 clone\_end:T7  
 vector\_side:right  
 misc\_feature 47673. .57070  
 /note="assembly\_name:Contig14"  
 misc\_feature 57171. .68217  
 /note="assembly\_name:Contig15"  
 misc\_feature 68318. .79880  
 /note="assembly\_name:Contig16"  
 misc\_feature 79981. .95382  
 /note="assembly\_name:Contig17"  
 misc\_feature 95483. .115764  
 /note="assembly\_name:Contig18"  
 misc\_feature 115865. .141797  
 /note="assembly\_name:Contig19"  
 misc\_feature 141898. .165593  
 /note="assembly\_name:Contig20"

## ORIGIN

Query Match 38.1%; Score 170; DB 2; Length 165593;  
 Best Local Similarity 80.0%; Pred. No. 5.5e-30;  
 Matches 200; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
 QY 1 GTAATGCAGATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGC 60  
 Db 74275 GCAGGGGCGGAATCTTGGCTCACTGCAAGCTCCAGCCTCCCGAGTAGCTGGGATTACAGGC 74334  
 QY 61 ATCGGCCACACGCCCGCGTAATTTTGTATCTTTTAGTAGAGCGGCTTCTCCATGTT 120  
 Db 74335 ATGTCACCATCTGCTGCTAAATTTTGTATTTTAGTAGAGATGGGGTTTCAACCTGTT 74394  
 QY 121 GGTGAGGCTGGTCTCGAACTTCAACCTCAGTGATCCGCCCGCTCGGCTCCCAAGT 180

Db 74395 GGTGAGGCTGGTCTCGAGCTCTTGACATCAGGTGATCTGCCCTCTCGGCTCCCAAGT 74454  
 QY 181 GCTAGGATTACAGGCGTGAGCCACCGGCTCAGCCTGGGAACACCTTTTCTTACATCTTC 240  
 Db 74455 GCTAGGATTACAGGCTTGAGCCACCGCGCCGCTCTTAAACCATCTTCTTGAGCCTCTCC 74514  
 QY 241 AAGTGCTAGA 250  
 Db 74515 CTGTGACTGA 74524

Search completed: January 15, 2005, 17:44:09  
 Job time : 2271 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 15, 2005, 14:58:47 ; Search time 355 Seconds  
(without alignments)  
6595.047 Million cell updates/sec

Title: US-10-009-579-5\_COPY\_3115\_3560

Perfect score: 446

Sequence: 1 gtaatggcagcatctctgct.....ctggaaggtctctgctgt 446

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- N\_Geneseq\_23Sep04:\*
- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	446	100.0	4282	4 AAL41898	Aal41898 Human GA7
2	170	38.1	568	4 AAL10010	Aal10010 Human CDN
3	169.6	38.0	4388	4 AAL04959	Aal04959 Human rep
4	169.6	38.0	4388	4 ABL97853	Ab197853 Human tes
5	169.2	37.9	3252	11 ADM01919	Adm01919 Human CDN
6	169.2	37.9	98800	12 ADN06353_3	Continuation (4 of
7	166.8	37.4	344	4 AAK68705	Aak68705 Human imm
8	166	37.2	26874	9 AAD56105	Aad56105 Human CCN
9	166	37.2	26874	9 ADA02467	Ada02467 Human CCN
10	166	37.2	26874	10 ADB72206	Adb72206 Human CCN
11	165.2	37.0	22651	4 AAK78202	Aak78202 Human imm
12	163.4	36.6	96596	10 ADE95968	Ade95968 Human NFA
13	163.4	36.6	96597	9 ADA02720	Ada02720 Human NFA
14	163.4	36.6	96597	10 ADB72458	Adb72458 Human NFA
15	162.8	36.5	87980	8 AAD53223_4	Continuation (5 of
16	162.8	36.5	247509	10 ADL13684	Adl13684 Osteoarth
17	162.4	36.4	185371	6 ABL10718	Abt10718 Human bre
18	162.2	36.4	3200	6 AAK47724	Aak47724 Genomic s
19	162.2	36.4	160361	12 ADL08116	Adl08116 Human gen
20	161.2	36.1	4779	4 AAI61042	Aai61042 Human pol
21	161	36.1	301	10 ADH59595	Adh59595 Alu-repea

22	160.2	35.9	1250	5 ADL62879	Adl62879 Human ova
23	160.2	35.9	7831	5 ABA19111	Abal9111 Human ner
24	160.2	35.9	23075	5 ABA19112	Abal9112 Human ner
25	160	35.9	382	3 AAC28799	Aac28799 Human sec
26	160	35.9	3062	4 AAI14445	Aai14445 Human CDN
27	160	35.9	3062	6 ABX04191	Abx04191 Human mRN
28	160	35.9	4990	4 AAK65261	Aak65261 Human imm
29	160	35.9	6534	4 AAL02766	Aal02766 Human rep
30	160	35.9	10324	4 AAK65263	Aak65263 Human imm
31	160	35.9	14136	4 AAK65264	Aak65264 Human imm
32	160	35.9	14387	4 AAK65265	Aak65265 Human imm
33	159.8	35.8	1744	6 ABZ70301	Abz70301 Human tyr
34	159.6	35.8	3656	10 ADB62811	Adb62811 Human CDN
35	159.6	35.8	5335	4 AAK69230	Aak69230 Human imm
36	159.6	35.8	6670	4 AAL02846	Aal02846 Human rep
37	159.6	35.8	7713	4 AAL02845	Aal02845 Human rep
38	159.4	35.7	294	5 ABA16264	Abal6264 Human ner
39	159.4	35.7	304	5 ABA16263	Abal6263 Human ner
40	159.4	35.7	645	4 AAK85882	Aak85882 Human imm
41	159.4	35.7	645	4 AAK85880	Aak85880 Human imm
42	159.4	35.7	10159	4 AAK73470	Aak73470 Human imm
43	159.4	35.7	10159	4 AAK73471	Aak73471 Human imm
44	159.2	35.7	19300	12 ADP74371	Adp74371 Human X c
45	159.2	35.7	91352	12 ADN94799	Adn94799 DNA encod

ALIGNMENTS

RESULT 1  
AAL41898

ID AAL41898 standard; DNA; 4282 BP.

XX AC AAL41898;

XX AC 03-MAY-2002 (first entry)

XX DE Human GA733-2 gene (encoding human epithelial glycoprotein-2) promoter.  
XX KW Human; GA733-2 gene promoter; gene; epithelial glycoprotein-2; EGP-2;  
XX KW pan-carcinoma associated antigen; cancer; carcinoma selective expression;  
XX KW treatment evaluation; non-squamous epithelium disease; carcinogenesis;  
XX KW transgenic animal; ds; gene therapy.

XX OS Homo sapiens.

XX PN EF1130106-A1.

XX XX 05-SEP-2001.

XX XX 01-MAR-2000; 2000EP-00200728.

XX XX 01-MAR-2000; 2000EP-00200728.

XX XX (UYGR-) RIJKSUNIV GRONINGEN.

XX PI De Leij LFMH, McLaughlin PMJ, Ruiter WHJ, Harmsen MC;

XX PI Van Der Molen H, Terpstra P, Dokter WHA,

XX WPI; 2001-591523/67.

XX DR Novel isolated and/or recombinant nucleic acid having tissue specific  
XX PT promoter derived from epithelial glycoprotein 2 gene, that allows  
XX PT expression of desired nucleic acid in cancer cell, specifically carcinoma  
XX PT cell.

XX PS Claim 3; Fig 1; 21pp; English.

XX CC The invention comprises the promoter sequence of the human GA733-2 gene.  
XX CC The GA733-2 gene encodes human epithelial glycoprotein-2 (EGP-2), which  
XX CC is a pan-carcinoma associated antigen. The GA733-2 gene promoter allows  
XX CC the expression of a nucleic acid of interest in a cancer cell, where the  
XX CC expression within the cancer cell is carcinoma selective. The GA733-2



KW Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX

PN W0200155320-A2.

XX  
PD 02-AUG-2001.

XX  
PF 17-JAN-2001; 2001WO-US001339.

XX  
PR 31-JAN-2000; 2000US-0179065P.

PR  
PR 04-FEB-2000; 2000US-0180628P.

PR  
PR 24-FEB-2000; 2000US-0184664P.

PR  
PR 02-MAR-2000; 2000US-0186350P.

PR  
PR 16-MAR-2000; 2000US-0189874P.

PR  
PR 17-MAR-2000; 2000US-0190076P.

PR  
PR 18-APR-2000; 2000US-0198123P.

PR  
PR 19-MAY-2000; 2000US-0205515P.

PR  
PR 07-JUN-2000; 2000US-0209467P.

PR  
PR 28-JUN-2000; 2000US-0214886P.

PR  
PR 30-JUN-2000; 2000US-0215135P.

PR  
PR 07-JUL-2000; 2000US-0216647P.

PR  
PR 07-JUL-2000; 2000US-0216880P.

PR  
PR 11-JUL-2000; 2000US-0217487P.

PR  
PR 11-JUL-2000; 2000US-0217496P.

PR  
PR 14-JUL-2000; 2000US-0218290P.

PR  
PR 26-JUL-2000; 2000US-0220963P.

PR  
PR 26-JUL-2000; 2000US-0220964P.

PR  
PR 14-AUG-2000; 2000US-0224518P.

PR  
PR 14-AUG-2000; 2000US-0224519P.

PR  
PR 14-AUG-2000; 2000US-0225213P.

PR  
PR 14-AUG-2000; 2000US-0225214P.

PR  
PR 14-AUG-2000; 2000US-0225266P.

PR  
PR 14-AUG-2000; 2000US-0225267P.

PR  
PR 14-AUG-2000; 2000US-0225268P.

PR  
PR 14-AUG-2000; 2000US-0225270P.

PR  
PR 14-AUG-2000; 2000US-0225447P.

PR  
PR 14-AUG-2000; 2000US-0225575P.

PR  
PR 14-AUG-2000; 2000US-0225758P.

PR  
PR 14-AUG-2000; 2000US-0225759P.

PR  
PR 18-AUG-2000; 2000US-0226279P.

PR  
PR 22-AUG-2000; 2000US-0226681P.

PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239352P.  
PR 13-OCT-2000; 2000US-0239357P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0241826P.  
PR 08-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 11-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.

```
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 7647; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 4388 BP; 870 A; 1238 C; 1205 G; 1075 T; 0 U; 0 Other;

Query Match 38.0%; Score 169.6; DB 4; Length 4388;
Best Local Similarity 86.6%; Pred. No. 1.4e-28;
Matches 187; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 GTAATGCCAGATCTCTGCTCACTGCAACCTCAGCTCCCGCTAGCTGGGATTACAGGC 60
Db |||||
QY 61 ATGCGCCACCACCGCGCTAAATTTGTATCTTTTAGTAGAGACGGCGTTCCTCCCATGTT 120
Db |||||
QY 121 GGTGAGCTGGTCTCGAACTTCAAACTCAGGTGATCCCGCGCTCGGCTCCCAAAGT 180
Db |||||
QY 1199 GCGCAGCTGGTCTCGAACTCCTGACCTCAAGTGATCCGCGCTCGGCTCCCAAAGT 1258
Db |||||
QY 181 GCTAGGATTACAGCGTGGAGCCACCGGCTCAGCCT 216
Db |||||
QY 1259 GCTGGAATTACAGCTTGAGTCACTGCGCGCGCCT 1294
Db |||||

RESULT 4
ABL97853
ID ABL97853 standard; DNA; 4388 BP.
XX
AC ABL97853;
XX
DT 21-JUN-2002 (first entry)
XX
DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2505.
XX
KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
FN WO20015317-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001329.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
```



SQ Sequence 3252 BP; 817 A; 802 C; 835 G; 798 T; 0 U; 0 Other;

Query Match 37.9%; Score 169.2; DB 11; Length 3252;  
Best Local Similarity 81.9%; Pred. No. 1.6e-29;  
Matches 195; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 1 GTAATGCGAGATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGC 60  
Db 2742 GCAGTGGCATGATCTCAGCTCACTGCAACCTCAGCCTCTCGAGTAGCTGGGATTACAGGC 2801

Qy 61 ATGGCCACACAGCCCGGCTAAATTTGTATCTTTTAGTAGAGACGGGHTTCTCCATGTT 120  
Db 2802 ACGTACCACACAGCCCTGGCTAAATTTGTATTTATTAGTAGAGACGGGHTTCAACCATGTT 2861

Qy 121 GGTCAAGCTGTCTCGAACTTCAAACTCAGGTGATCGGCCCGCTCGGCTCCCAAGT 180  
Db 2862 GGCAGGCTGGTCTCAAACTCCTGACCTCAGTGATCCACTCCCTCAGCCTCCCAAGT 2921

Qy 181 GCTAGGATTACAGCGGTGAGCCACCGGCTCAGCCTGGGAACACCTTTTCTTACATCT 238  
Db 2922 ACTGGGATTACAGCGGTGAGCCACTGGCCCACTGAGAAGCAGCTTTAAATACATTT 2979

RESULT 6  
ADN06353\_3/c  
Continuation (4 of 4) of ADN06353 from base 300001 (Human FLAP genomic DNA SEQ ID NO:1.  
WP Sequence split into 4 fragments LOCUS ADN06353 Accession Adn06353  
WP Fragment Name Begin End  
WP ADN06353\_0 1 110000  
WP ADN06353\_1 100001 210000  
WP ADN06353\_2 200001 310000  
WP ADN06353\_3 300001 398800

Query Match 37.9%; Score 169.2; DB 12; Length 98800;  
Best Local Similarity 85.1%; Pred. No. 3.3e-28;  
Matches 189; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 23 CTGCAACTCAGCCTCCCGAGTAGTGGGATTACAGGCATGCGCCACCGCCGCGCTAA 82  
Db 81352 CTCTGCTCAGCCTCCCGAGTAGTGGGATTACAGGCATGCAACACCGTCTGGCTGA 81293

Qy 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTTGGTCAAGCTGTTCTCGAACTTC 142  
Db 81292 TTTTCTATTTTTAGTAGAGATGGGTTTCTCCATGTTGGTCAAGCTGTTCTCGAACTCC 81233

Qy 143 AAACCTCAGTGATCGCCCGCTCGGCTCCCAAGTGTAGATTACAGGCGTGAGCC 202  
Db 81232 TGACCTCAGTGATCGCCCGCTCGGCTCCCAAGTGTGGGATTACAGGCGTGAGCC 81173

Qy 203 ACCGCGCTCAGCCTGGGAACACCTTTTCTTACATCTTCAAGT 244  
Db 81172 ACCACGCCAGCCAAGCAATTCATTTCAAAAGCTTTTAAT 81131

RESULT 7  
AAK68705/c  
ID AAK68705 standard; DNA; 344 Bp.  
XX  
AC AAK68705;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23517.  
XX  
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
PD  
XX

PF 17-JAN-2001; 2001WO-US001354.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0228279P.  
PR 22-AUG-2000; 2000US-0228681P.  
PR 22-AUG-2000; 2000US-0228688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0234080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.





PT New recombinant nucleic acid and its encoded protein, useful for  
 XX preparing a composition for diagnosing or treating carcinomas.  
 PS Claim 1; Page 103-107; 173pp; English.  
 CC  
 CC The invention relates to novel sequences which are useful for preparing a  
 CC composition for diagnosing or treating carcinomas. These sequence are  
 CC also useful in gene therapy. The present sequence is human CCND3  
 CC carcinoma associated (CA) gene. This sequence is used in the invention  
 XX  
 SQ Sequence 26874 BP; 6621 A; 6603 C; 6750 G; 6880 T; 0 U; 20 Other;  
 Query Match 37.2%; Score 166; DB 8; Length 26874;  
 Best Local Similarity 81.1%; Pred. No. 1.3e-27;  
 Matches 193; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
 QY 1 GTAATGCCAGATCTCTGCTCACTGCAACTTCAAACTCTCAGGTGATCCGCCGCTCCCAAGT 180  
 Db 22835 GCAGTGGCATGATCTCAGTCACTGCAACTCTCAGCTCTCGAGTAGCTGGGATTACAGGC 22776  
 QY 61 ATGCGCCACCCAGCCCGGCTAATTTGTATCTTTTAGTAGAGACGGGTTCCCTCATGTT 120  
 Db 22775 ACGTACCACCCAGCGCTCGGCTAATTTGTATTTAGTAGAGACAGGGTTTCACCATGTT 22716  
 QY 121 GGTCAAGCTGTCTCGAACTTCAAACTCTCAGGTGATCCGCCGCTCCCAAGT 180  
 Db 22715 GGCAGAGCTGGTCTCAAACTCTCGACCTCAGTGTACCTCTCCCTCAGCTCCCAAGT 22656  
 QY 181 GCTAGGATTACAGCGTGTAGCCACCGGCTCAGCTGTGGGAACACCTTTCTTACATCT 238  
 Db 22655 ACTGGGATTACAGCGTGTAGCCACTGGGCCCACTGTAGAGAGCGCTTTAAATACATTT 22598  
 RESULT 9  
 ADA02467/c  
 ID ADA02467 standard; DNA; 26874 BP.  
 XX  
 AC ADA02467;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE Human CCND3 carcinoma associated gene, SEQ ID NO:986.  
 XX  
 KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
 KW Gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003057146-A2.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 26-DEC-2002; 2002WO-US041414.  
 XX  
 PR 26-DEC-2001; 2001US-00035832.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX  
 PI Morris DW;  
 XX  
 XX WPI; 2003-587068/55.  
 DR  
 XX  
 PT New recombinant nucleic acid encoding carcinoma associated protein,  
 PT useful for preparing compositions for treating carcinomas.  
 XX  
 XX Claim 1; SEQ ID NO 986; 245pp; English.  
 PS  
 CC The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically

CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed human CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 26874 BP; 6621 A; 6602 C; 6751 G; 6880 T; 0 U; 20 Other;  
 Query Match 37.2%; Score 166; DB 9; Length 26874;  
 Best Local Similarity 81.1%; Pred. No. 1.3e-27;  
 Matches 193; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
 QY 1 GTAATGCCAGATCTCTGCTCACTGCAACTCTCAGCTCCCAAGTGGGATTACAGGC 60  
 Db 22835 GCAGTGGCATGATCTCAGTCACTGCAACTCTCAGCTCTCGAGTAGCTGGGATTACAGGC 22776  
 QY 61 ATGCGCCACCCAGCCCGGCTAATTTGTATCTTTTAGTAGAGACGGGTTCCCTCATGTT 120  
 Db 22775 ACGTACCACCCAGCGCTCGGCTAATTTGTATTTAGTAGAGACAGGGTTTCACCATGTT 22716  
 QY 121 GGTCAAGCTGTCTCGAACTTCAAACTCTCAGGTGATCCGCCGCTCCCAAGT 180  
 Db 22715 GGCAGAGCTGGTCTCAAACTCTCGACCTCAGTGTACCTCTCCCTCAGCTCCCAAGT 22656  
 QY 181 GCTAGGATTACAGCGTGTAGCCACCGGCTCAGCTGTGGGAACACCTTTCTTACATCT 238  
 Db 22655 ACTGGGATTACAGCGTGTAGCCACTGGGCCCACTGTAGAGAGCGCTTTAAATACATTT 22598  
 RESULT 10  
 ADB72206/c  
 ID ADB72206 standard; DNA; 26874 BP.  
 XX  
 AC ADB72206;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human CCND3 gene.  
 XX  
 KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
 KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003008583-A2.  
 XX  
 PD 30-JAN-2003.  
 XX  
 PF 26-DEC-2001; 2001WO-US051291.  
 XX  
 PR 02-MAR-2001; 2001US-00798586.  
 PR 23-OCT-2001; 2001US-00004113.  
 PR 08-NOV-2001; 2001US-00052482.  
 PR 30-NOV-2001; 2001US-00997722.  
 PR 20-DEC-2001; 2001US-00034650.  
 XX  
 XX (SAGR-) SAGRES DISCOVERY.  
 PA  
 XX Morris DW, Engelhard EK;  
 PI  
 XX WPI; 2003-239337/23.  
 DR

```
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
XX Claim 1; SEQ ID NO 34; 2304pp; English.
XX
XX The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a human gene of the invention.
XX
SQ Sequence 26874 BP; 6621 A; 6602 C; 6751 G; 6880 T; 0 U; 20 Other;

Query Match 37.2%; Score 166; DB 10; Length 26874;
Best Local Similarity 81.1%; Pred. No. 1.3e-27;
Matches 193; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1 GTAATGGCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTGGGATTACAGGC 60
Db 22835 GCAGTGGCAGCATCTCAGCTCACTGCAACCTCAGCCTCTCGAGTGGGATTACAGGC 22776

QY 61 ATGGCCACACACGCCCGCTAAATTTGTATCTTTTAGTAGAGCGGCTTCTCCATGTT 120
Db 22775 ACGTACCACACACGCCCTGGCTAAATTTTGTATTTATTAGTAGAGCAGGTTTCACCATGTT 22716

QY 121 GGTGAGCTGCTCTGAACTTCAAACTCAGGTGATCCGCGGCTCGGCTCCCAAAGT 180
Db 22715 GGCAGGCTGCTCTCAAACTCCTGACCTCAGTGCATCCCTCAGCTCCCAAAGT 22656

QY 181 GCTAGGATTACAGCGTGAGCCACCGGCTCAGCTCGGACACCTTTCTTACATCT 238
Db 22655 ACTGGGATTACAGCGTGAGCCACTGCGCCCACTGAGAGCAGCTTTAAATACATTT 22598

RESULT 11
AAK78202/c
ID AAK78202 standard; DNA; 22651 BP.
XX
AC AAK78202;
XX
XX 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33014.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX
XX 18-APR-2000; 2000US-0198123P.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX 28-JUN-2000; 2000US-0214886P.
XX
XX 30-JUN-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216647P.
XX
XX 11-JUL-2000; 2000US-0216880P.
XX
XX 11-JUL-2000; 2000US-0217487P.
XX
XX 01-NOV-2000; 2000US-0217496P.
XX
XX 14-JUL-2000; 2000US-0218290P.
XX
XX 26-JUL-2000; 2000US-0220963P.
XX
XX 26-JUL-2000; 2000US-0220964P.
XX
XX 14-AUG-2000; 2000US-0224518P.
XX
XX 14-AUG-2000; 2000US-0224519P.
XX
XX 14-AUG-2000; 2000US-0225213P.
XX
XX 14-AUG-2000; 2000US-0225214P.
XX
XX 14-AUG-2000; 2000US-0225266P.
XX
XX 14-AUG-2000; 2000US-0225267P.
XX
XX 14-AUG-2000; 2000US-0225268P.
XX
XX 14-AUG-2000; 2000US-0225270P.
XX
XX 14-AUG-2000; 2000US-0225447P.
XX
XX 14-AUG-2000; 2000US-0225757P.
XX
XX 14-AUG-2000; 2000US-0225758P.
XX
XX 14-AUG-2000; 2000US-0225759P.
XX
XX 18-AUG-2000; 2000US-0226279P.
XX
XX 22-AUG-2000; 2000US-0226681P.
XX
XX 22-AUG-2000; 2000US-0226686P.
XX
XX 22-AUG-2000; 2000US-0227182P.
XX
XX 23-AUG-2000; 2000US-0227009P.
XX
XX 30-AUG-2000; 2000US-0228924P.
XX
XX 01-SEP-2000; 2000US-0229287P.
XX
XX 01-SEP-2000; 2000US-0229343P.
XX
XX 01-SEP-2000; 2000US-0229344P.
XX
XX 01-SEP-2000; 2000US-0229345P.
XX
XX 05-SEP-2000; 2000US-0229509P.
XX
XX 05-SEP-2000; 2000US-0229513P.
XX
XX 06-SEP-2000; 2000US-0230437P.
XX
XX 06-SEP-2000; 2000US-0230438P.
XX
XX 08-SEP-2000; 2000US-0231242P.
XX
XX 08-SEP-2000; 2000US-0231243P.
XX
XX 08-SEP-2000; 2000US-0231244P.
XX
XX 08-SEP-2000; 2000US-0231413P.
XX
XX 08-SEP-2000; 2000US-0231414P.
XX
XX 08-SEP-2000; 2000US-0232080P.
XX
XX 08-SEP-2000; 2000US-0232081P.
XX
XX 12-SEP-2000; 2000US-0231968P.
XX
XX 14-SEP-2000; 2000US-0232397P.
XX
XX 14-SEP-2000; 2000US-0232398P.
XX
XX 14-SEP-2000; 2000US-0233063P.
XX
XX 14-SEP-2000; 2000US-0233064P.
XX
XX 14-SEP-2000; 2000US-0233065P.
XX
XX 21-SEP-2000; 2000US-0234223P.
XX
XX 21-SEP-2000; 2000US-0234274P.
XX
XX 25-SEP-2000; 2000US-0234997P.
XX
XX 25-SEP-2000; 2000US-0234998P.
XX
XX 26-SEP-2000; 2000US-0235484P.
XX
XX 27-SEP-2000; 2000US-0235834P.
XX
XX 27-SEP-2000; 2000US-0235836P.
XX
XX 29-SEP-2000; 2000US-0236327P.
XX
XX 29-SEP-2000; 2000US-0236367P.
XX
XX 29-SEP-2000; 2000US-0236368P.
XX
XX 29-SEP-2000; 2000US-0236370P.
XX
XX 02-OCT-2000; 2000US-0236802P.
XX
XX 02-OCT-2000; 2000US-0237037P.
XX
XX 02-OCT-2000; 2000US-0237038P.
XX
XX 02-OCT-2000; 2000US-0237039P.
XX
XX 02-OCT-2000; 2000US-0237040P.
XX
XX 13-OCT-2000; 2000US-0239355P.
XX
XX 13-OCT-2000; 2000US-0239377P.
XX
XX 20-OCT-2000; 2000US-0240960P.
XX
XX 20-OCT-2000; 2000US-0241221P.
XX
XX 20-OCT-2000; 2000US-0241785P.
XX
XX 20-OCT-2000; 2000US-0241786P.
XX
XX 20-OCT-2000; 2000US-0241787P.
XX
XX 20-OCT-2000; 2000US-0241808P.
XX
XX 20-OCT-2000; 2000US-0241809P.
XX
XX 20-OCT-2000; 2000US-0241826P.
XX
XX 01-NOV-2000; 2000US-0244617P.
```

PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 08-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX Disclosure; SEQ ID NO 33014; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patient's own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting the  
XX nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/hematopoietic-related diseases, especially

CC cancers and cancer metaataees of haematopoietic-derived cells, AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention  
XX  
XX  
SQ Sequence 22651 BP; 5939 A; 5107 C; 5360 G; 6245 T; 0 U; 0 Other;  
  
Query Match 37.0%; Score 165.2; DB 4; Length 22651;  
Best Local Similarity 90.7%; Pred. No. 1.9e-27;  
Matches 176; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
  
QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGATGCGCCACACGCCCGGCTAA 82  
DB 4434 CTCCTGCCCTCAGCCTCCCGAGTAGCTGGGATTACAGGATGCGCCACACGCCCGGCTAA 4375  
QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTTGGTCAGGCTGTCTCGAACTTC 142  
DB 4374 TTTTGTATCTTTTAGTAGAGATGGGGTTTCTCCATGTTGGTCAGGCTGTCTCGAACTTC 4315  
QY 143 AAACCTCAGGTGATCGCGCCGCTCGGCTCCCAAGTCTAGGATTACAGGCGTGAGCC 202  
DB 4314 CAATCTCAGGTGATCGCGCCGCTCAGTCTCCAAAGTCTGGAATTACAGGCGTGAGCC 4255  
QY 203 ACCGGGCTCAGCCT 216  
DB 4254 ACGGCACCCAGCCT 4241  
  
RESULT 12  
ADE95968  
ID ADE95968 standard; DNA; 96596 BP.  
XX  
AC ADE95968;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Human NFATC1 gene genomic DNA sequence.  
XX  
KW cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;  
XX lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; NFATC1.  
XX Homo sapiens.  
XX  
XX WO2003039484-A2.  
XX  
XX 15-MAY-2003.  
XX  
XX 08-NOV-2002; 2002WO-US036071.  
XX  
XX 08-NOV-2001; 2001US-00052482.  
XX  
XX (SAGR-) SAGRES DISCOVERY.  
XX  
XX Morris DW, Engelhard EK;  
XX  
XX WPI; 2003-441462/41.  
XX  
XX New carcinoma associated nucleic acids and proteins, useful for screening  
XX drug candidates, or for diagnosing and treating carcinomas, e.g.  
XX lymphoma, breast cancer, prostate cancer or leukemia.  
XX  
XX Claim 1; SEQ ID NO 226; 793pp; English.  
XX  
XX This invention relates to novel recombinant nucleic acids for use in  
XX diagnosis and treatment of cancer, especially carcinomas, as well as the  
XX use of compositions in screening methods. The compositions of the  
XX invention may have cytostatic activity whilst the disclosed sequences may  
XX be useful for gene therapy. The carcinoma associated nucleic acids and  
XX proteins are useful for diagnosing and treating carcinomas, for example  
XX lymphoma, breast cancer, prostate cancer or leukaemia, or for screening  
XX drug candidates or bioactive agents capable of binding to, or modulating  
XX the activity of, a carcinoma associated protein. The present sequence is  
XX the genomic DNA sequence of the human NFATC1 gene which is a carcinoma

CC associated gene of the invention.  
XX  
SQ Sequence 96596 BP; 17605 A; 24295 C; 25661 G; 19918 T; 0 U; 9117 Other;  
  
Query Match 36.6%; Score 163.4; DB 10; Length 96596;  
Best Local Similarity 87.6%; Pred. No. 6.6e-27;  
Matches 190; Conservative 0; Mismatches 26; Indels 1; Gaps 1;  
  
QY 1 GTAATGGCAGATCTCTGCTCACTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGC 60  
Db 76410 GCAGTGGCGTGATCTCGGCTCACTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGC 76469  
  
QY 61 ATGCGCCACACGCGCGCTAAATTTGTATCTTTTAGTAGAGAGCGGTTCTCCCATGTT 120  
Db 76470 ATGACACACACGCGCGCTAAATTTGTATCTTTTAGTAGAGAGTAGGGTTTTCGCATGTT 76528  
  
QY 121 GGTGAGCTGGTCTCGAATCTCAAACTCAGGTGATCCGCGCTCGGCTCCCAAAGT 180  
Db 76529 GGCCAAGCAGGTCTGGAACTCCCAACCTCAGGTGATCCGCGCTCGGCTCCCAAAGT 76588  
  
QY 181 GCTAGGATTACAGCGTGAGCCACCGCGCTCAGGCTG 217  
Db 76589 GCTGGGATTACAGCATGAGCCACCGTGCCCGGCTG 76625  
  
RESULT 13  
ADA02720 ID ADA02720 standard; DNA; 96597 BP.  
AC ADA02720;  
XX  
XX 06-NOV-2003 (first entry)  
XX Human NFATC1 carcinoma associated gene, SEQ ID NO:1238.  
DE  
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO2003057146-A2.  
XX  
XX 17-JUL-2003.  
XX  
XX 26-DEC-2002; 2002WO-US041414.  
XX  
XX 26-DEC-2001; 2001US-00035832.  
XX  
XX (SAGR-) SAGRES DISCOVERY.  
XX  
XX Morris DW;  
XX  
XX WPI; 2003-587068/55.  
XX  
XX New recombinant nucleic acid encoding carcinoma associated protein,  
PT useful for preparing compositions for treating carcinomas.  
XX  
XX Claim 1; SEQ ID NO 1238; 245pp; English.  
XX  
XX The invention relates to recombinant carcinoma associated (CA) nucleic  
CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
CC invention also encompasses expression vectors and host cells comprising a  
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
CC binds to the protein, and a biochip comprising CA nucleic acid or  
CC fragments thereof. The sequences of the invention were identified using  
CC oncogenic retroviruses, which insert into the genome of the host organism  
CC at random. Many of these do not carry transduced host oncogenes or  
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
CC direct consequence of the effects of proviral integration into host  
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or

CC leukaemia) or a propensity to carcinoma by determination of the sequence  
CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed human CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 96597 BP; 17605 A; 24295 C; 25662 G; 19918 T; 0 U; 9117 Other;  
  
Query Match 36.6%; Score 163.4; DB 9; Length 96597;  
Best Local Similarity 87.6%; Pred. No. 6.6e-27;  
Matches 190; Conservative 0; Mismatches 26; Indels 1; Gaps 1;  
  
QY 1 GTAATGGCAGATCTCTGCTCACTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGC 60  
Db 76411 GCAGTGGCGTGATCTCGGCTCACTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGC 76470  
  
QY 61 ATGCGCCACACGCGCGCTAAATTTGTATCTTTTAGTAGAGAGCGGTTCTCCCATGTT 120  
Db 76471 ATGACACACACGCGCGCTAAATTTGTATCTTTTAGTAGAGAGTTTTCGCATGTT 76529  
  
QY 121 GGTGAGCTGGTCTCGAATCTCAAACTCAGGTGATCCGCGCTCGGCTCCCAAAGT 180  
Db 76530 GGCCAAGCAGGTCTGGAACTCCCAACCTCAGGTGATCCGCGCTCGGCTCCCAAAGT 76589  
  
QY 181 GCTAGGATTACAGCGTGAGCCACCGCGCTCAGGCTG 217  
Db 76590 GCTGGGATTACAGCATGAGCCACCGTGCCCGGCTG 76626  
  
RESULT 14  
ADB72458 ID ADB72458 standard; DNA; 96597 BP.  
XX  
XX ADB72458;  
XX  
XX 04-DEC-2003 (first entry)  
XX Human NFATC1 gene.  
XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.  
XX  
OS Homo sapiens.  
XX  
XX WO2003008583-A2.  
XX  
XX 30-JAN-2003.  
XX  
XX 26-DEC-2001; 2001WO-US051291.  
XX  
XX 02-MAR-2001; 2001US-00798586.  
XX  
XX 23-OCT-2001; 2001US-00004113.  
XX  
XX 08-NOV-2001; 2001US-00052482.  
XX  
XX 30-NOV-2001; 2001US-00997722.  
XX  
XX 20-DEC-2001; 2001US-00034650.  
XX  
XX (SAGR-) SAGRES DISCOVERY.  
XX  
XX Morris DW, Engelhard EK;  
XX  
XX WPI; 2003-239337/23.  
XX  
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
PT cancers, neoplasm, adenocarcinoma, or sarcomas.  
XX  
XX Claim 1; SEQ ID NO 286; 2304pp; English.  
XX  
XX The invention relates to a novel recombinant nucleic acid comprising a  
CC nucleotide sequence selected from any of the 660 sequences fully defined

CC in the specification. A polynucleotide of the invention has cytostatic  
CC activity, and may have a use in gene therapy, or in a vaccine. The  
CC recombinant nucleic acids and polypeptides are useful for treating  
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
CC sarcomas. The present sequence represents a human gene of the invention.

XX  
SQ Sequence 96597 BP; 17605 A; 24295 C; 25662 G; 19918 T; 0 U; 9117 Other;

Query Match 36.6%; Score 163.4; DB 10; Length 96597;  
Best Local Similarity 87.6%; Pred. No. 6.6e-27;  
Matches 190; Conservative 0; Mismatches 26; Indels 1; Gaps 1;  
  
QY 1 GTAATGCGACGATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGC 60  
DB 76411 GCAGTGGCGTGATCTCGGCTCACTGCAACCTCCACCTCCCGAGTAGCTGGGATTACAGGC 76470  
  
QY 61 ATGGGCGACACCGCGGGCTAATTTGTATCTTTTAGTAGAGACGGCGTTCCTCCATGTT 120  
DB 76471 ATGCACCGACACCGCGGGCTAATTTGTATCTTTTAGTAGAGATGGGTTTGGCCATGTT 76529  
  
QY 121 GGTGAGCTGGTCTCGAAGCTTCAACCTCAGGTGATCCGCGCTCGGCTCCCAAGT 180  
DB 76530 GGCACAGCAGGTCTGGAAGCTTCAACCTCAGGTGATCCGCGCTCGGCTCCCAAGT 76589  
  
QY 181 GCTAGGATTACAGCGCTGAGCCACCGCGCTCAGCCTG 217  
DB 76590 GCTGGGATTACAGCGCATGAGCCACCGTGCCGCGCTG 76626

## RESULT 15

AAD53223\_4  
Continuation (5 of 5) of AAD53223 from base 400001 (Human chromosome 3 p-arm breakpoint  
WP Sequence split into 5 fragments LOCUS AAD53223 Accession AAD53223

WP	Fragment Name	Begin	End
WP	AAD53223_0	1	110000
WP	AAD53223_1	100001	210000
WP	AAD53223_2	200001	310000
WP	AAD53223_3	300001	410000
WP	AAD53223_4	400001	487980

Query Match 36.5%; Score 162.8; DB 8; Length 87980;  
Best Local Similarity 88.9%; Pred. No. 8.9e-27;  
Matches 176; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGCATGCGCCACCGCCGCTAA 82  
DB 33281 CTCTACCTCAGCCTCCCGAGTAGCTGGGATTACAGCATGCGCCACCGCTGCTAA 33340  
  
QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCCTCCATGTTGGTTCAGGCTGGTCTCGAACTTC 142  
DB 33341 TTTTGTATCTTTTAGTAGAGACGAGGTTTCTCCATGTTGGTTCAGGCTGGTCTCGAACTTC 33400  
  
QY 143 AAACCTCAGGTGATCGCGCGCTCGGCTCCCAAGTGTAGGATTACAGCGGTGAGCC 202  
DB 33401 TGACCTTATGTGATCGCGCGCTCGGCTCCCAAGTGTAGGATTACAGCGGTGAGCC 33460  
  
QY 203 ACCGCGCTCAGCCTGGGA 220  
DB 33461 AATGCGCCGACGCGAGGA 33478

Search completed: January 15, 2005, 17:06:20  
Job time : 364 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 15, 2005, 16:47:08 ; Search time 74 Seconds  
(without alignments)  
4283.944 Million cell updates/sec

Title: US-10-009-579-5\_COPY\_3115\_3560

Perfect score: 446

Sequence: 1 gtaatggcagatctctgct.....ctggaaggtctctgctgt 446

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 35539441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	162.2	36.4	8220	4	US-09-797-908-3
C 2	160	35.9	382	4	US-09-513-999C-32874
C 3	158.8	35.6	84495	3	US-09-797-906-3
C 4	158	35.4	462	4	US-09-621-976-14372
C 5	157.4	35.3	70000	4	US-09-851-896-3
C 6	156.4	35.1	1811	1	US-08-848-252-1
C 7	154	34.5	63000	4	US-09-780-172-18
C 8	153.4	34.4	354	4	US-09-621-976-8798
C 9	153.4	34.4	373	4	US-09-621-976-14426
C 10	153.4	34.4	841	5	PCT-US93-06251-80
C 11	153.4	34.4	841	5	PCT-US93-06251-81
C 12	153.4	34.4	66933	4	US-09-544-398B-11
C 13	153.4	34.4	66933	4	US-09-543-771-11
C 14	153.4	34.4	72049	4	US-09-544-398B-9
C 15	153.4	34.4	72049	4	US-09-543-771-9
C 16	153.2	34.3	609	3	US-09-385-982-237
C 17	153.2	34.3	1371	4	US-09-023-655-986
C 18	153.2	34.3	31571	1	US-08-323-443B-1
C 19	153.2	34.3	53526	3	US-08-658-136-2
C 20	153.2	34.3	53577	3	US-08-658-136-1
C 21	153	34.3	282	1	US-08-133-629-8
C 22	153	34.3	283	4	US-08-579-445-26
C 23	152.8	34.3	2839	3	US-09-061-702-1
C 24	152.8	34.3	2839	4	US-09-748-451-1
C 25	152.8	34.3	168174	4	US-10-071-411A-63
C 26	152.8	34.3	168273	4	US-10-071-411A-2
C 27	151.8	34.0	21721	4	US-09-269-939A-41

28	151.6	34.0	50000	3	US-09-146-053-3	Sequence 3, Appli
C 29	151.4	33.9	4233	3	US-09-056-105-27	Sequence 27, Appli
C 30	151.4	33.9	26016	4	US-09-326-480A-1	Sequence 1, Appli
31	151.4	33.9	55298	4	US-09-491-356C-1	Sequence 1, Appli
32	150.6	33.8	8174	1	US-07-914-281-5	Sequence 5, Appli
33	150.6	33.8	8174	1	US-08-393-246-5	Sequence 5, Appli
34	150.6	33.8	8174	1	US-08-525-058A-5	Sequence 5, Appli
35	150.6	33.8	8174	2	US-08-696-731-5	Sequence 5, Appli
36	150.6	33.8	8174	3	US-09-042-531-5	Sequence 5, Appli
37	150.6	33.8	8174	5	PCT-US91-00899-3	Sequence 8, Appli
38	150.2	33.7	9365	3	US-09-508-285A-8	Sequence 8, Appli
39	150.2	33.7	9365	3	US-09-350-836B-8	Sequence 8, Appli
40	150.2	33.7	9365	4	US-09-370-265-8	Sequence 8, Appli
41	150.2	33.7	9365	4	US-09-557-800C-8	Sequence 8, Appli
42	150.2	33.7	9365	4	US-09-370-625A-8	Sequence 8, Appli
43	150.2	33.7	14747	3	US-09-608-285A-42	Sequence 42, Appli
44	150.2	33.7	14747	4	US-09-557-800C-42	Sequence 42, Appli
45	150.2	33.7	15977	3	US-09-608-285A-59	Sequence 59, Appli

ALIGNMENTS

RESULT 1

US-09-797-908-3/c

; Sequence 3, Application US/09797908

; Patent No. 6555352

; GENERAL INFORMATION:

; APPLICANT: WEBSTER, Marion et al.

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; FILE REFERENCE: CL00781

; CURRENT APPLICATION NUMBER: US/09/797,908

; CURRENT FILING DATE: 2001-03-05

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 8220

; TYPE: DNA

; ORGANISM: Human

; US-09-797-908-3

Query Match 36.4%; Score 162.2; DB 4; Length 8220;  
Best Local Similarity 90.6%; Pred. No. 1.9e-32;  
Matches 173; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy	23	CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCCGGCTAA	82
Db	748	CTCCCGCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCCGGCTAA	689
Qy	83	TTTTGTATCTTTTAGTAGAGACGGCTTCCTCCATGTTGTCAGGCTGTTCTGAACCTC	142
Db	688	TTTTGTATTTTTTAGTAGAGACGGGTTTCTCCATGTTGTCAGGCTGTTCTGAACCTC	629
Qy	143	AAACCTCAGTGATCGCCCGCTCGCCCTCCCAAGTGTAGGATTACAGGCGGTGAGCC	202
Db	628	CAACCTCAGTGATCTGCGCTCGCTCCAGCTCCCAAGTGTGAGATTACAGCGTGAGCC	569
Qy	203	ACCGCGCTCAG 213	
Db	568	ACCGCGCCAG 558	

RESULT 2

US-09-513-999C-32874/c

; Sequence 32874, Application US/09513999C

; Patent No. 6783961

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

```
; Patent No. 6783961
; FILE REFERENCE: 59, US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 32874
; LENGTH: 382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 283
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 284
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 311
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 319
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 321
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 352
; OTHER INFORMATION: s=g or c
; OTHER INFORMATION: s=g or c
; US-09-513-999C-32874
```

```
Query Match 35.9%; Score 160; DB 4; Length 382;
Best Local Similarity 85.3%; Pred. No. 2.2e-32;
Matches 186; Conservative 2; Mismatches 29; Indels 1; Gaps 1;

QY 1 GTAATGCCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGC 60
Db 344 GTAATGCCAGCATCTGCGCTCACWMAACCTCKCCTCCGCGTAGCTGGATTACAGC 285

QY 61 ATGCCGACACGCCCGCGTAATTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTT 120
Db 284 NNGTGCACAAACGCCCACTAATTTTGTAT-TTTTAGTAGAGACAGGGTTTCTCCATATT 226

QY 121 GGTGAGGCTGCTCTCGAACTTCAAACTCAGGTGATCCGCCCGCTCCGCTCCCAAAGT 180
Db 225 GGTGAGGCTGCTCTCGAGCTCCGAGCTCAGGTGATCCGCCCACTCAGCCTCCCAAAGT 166

QY 181 GCTAGGATTACAGCGGTGAGCCACCGCGCTCAGCCTGG 218
Db 165 GCTGGGATTACAGCGGTGAGCCACCACTCTCTGCGCTGG 128
```

```
RESULT 3
US-09-797-906-3/c
; Sequence 3, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
; APPLICANT: Zhanghe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001151CIP
; CURRENT APPLICATION NUMBER: US/09/797,906
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 3
; LENGTH: 84495
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(84495)
; OTHER INFORMATION: n = A,T,C or G
; US-09-797-906-3

Query Match 35.6%; Score 158.8; DB 3; Length 84495;
Best Local Similarity 88.7%; Pred. No. 3.4e-31;
Matches 172; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCACACACGCCCGGCTAA 82
Db 46896 CTCCTGCCTCAGCTACCGAGTAGCTGGGATTAGAGGGCTGCGCCACCAACGCCCGGCTAA 46837

QY 83 TTTTGTATCTTTTAGTAGAGACGGGTTCTCTCATGTTGGTCAGCTGGTCTCGAACTTC 142
Db 46836 TTTTATATTTTATTAGTAGAGACGGGTTTCTCCATGTTGGTCAGCTGGTCTTTGAACTCC 46777

QY 143 AAACCTCAGGTGATCGCCCGCTCGGCCTCCCAAAGTCTAGGATTACAGCGGTGAGCC 202
Db 46776 CAACCTCAAGTGATCTGCGCGCTCGGCCTCTCAAAGTCTGCGGATTACAGGGGTGAGCC 46717

QY 203 ACCGGCTCAGCCT 216
Db 46716 ACCGGCGCGGCTT 46703
```

## RESULT 4

```
US-09-621-976-14372/c
; Sequence 14372, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 14372
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-14372
```

```
Query Match 35.4%; Score 158; DB 4; Length 462;
Best Local Similarity 67.8%; Pred. No. 7.9e-32;
Matches 236; Conservative 0; Mismatches 110; Indels 2; Gaps 1;

QY 29 CCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCACACCGCGCTAAATTTTGT 88
Db 369 CCTCAGCCTCCCGAGTAGCTGGGATTAGGCATGACACCAACCGCTGGCTAAATTTTG 310

QY 89 ATCTTTTAGTAGAGACGGGTTCTCTCCATGTTGGTCAGGCTGGTCTCGAACTTCAAACCT 148
Db 309 TATGTTTAGTAGAGATGGGTTCTACCAATGTTGGCCAGGCTGGTCTCAAACCTCTGACCT 250

QY 149 CAGGTGATCCCGCGCTCCCGAGTCTCCAAAGTGTAGGATTACAGCGGTGAGCCACCGG 208
Db 249 CAGGTGATCCCACTCCCTGCGTCCCAAAGTGTGGGATTACAGGAGTGAAGCCCGG 190

QY 209 CTCAGCCTGGGAACACCTTTCTTACATCTTCAAGTGTAGAAATGCTTATGAAACGAA 268
Db 189 CCCAGCCAAATATATATTTTAAATAATTCATGTTTAAATA--ATTATGAAAGTAAAC 132

QY 269 AAAAGAATTATTAAGAGTAATTTATAAAGAAACACTCATTTTCTCCCAAGAGAGCAAGA 328
```





; NUMBER OF SEQ ID NOS: 96  
; SEQ ID NO 18  
; LENGTH: 63000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-780-172-18

Query Match 34.5%; Score 154; DB 4; Length 63000;  
Best Local Similarity 87.4%; Pred. No. 5.4e-30;  
Matches 180; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

QY 23 CTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACGCGCGCTAA 82  
Db 34637 CTCCTGCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACGCGCTAA 34578  
QY 83 TTTTGTATCTTTTAGTAGAGACGGCTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142  
Db 34577 TTTTGTAT-TTTTAGTAGAGATAGGGTTTCTCTATGTTGGTCAGGCTGGTCTCAAACTCC 34519  
QY 143 AAACCTCAGGTGATCCCGCGCTCGCGCTCCCAAAGTGTAGATTACAGGCGTGAGCC 202  
Db 34518 CGACCTCAGGTGATCCCGCTCTCTCGCGCTCCCAAAGTGTAGATTACAGGCGTGAGCC 34459  
QY 203 ACCGCGCTCAGCTGGGAACACCTTT 228  
Db 34458 ACCTCGCCAGCCCAAGAAATATGTCT 34433

## RESULT 8

US-09-621-976-8798/c  
; Sequence 8798, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 8798  
; LENGTH: 354  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-8798

Query Match 34.4%; Score 153.4; DB 4; Length 354;  
Best Local Similarity 83.4%; Pred. No. 1.1e-30;  
Matches 186; Conservative 0; Mismatches 36; Indels 1; Gaps 1;  
QY 23 CTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACGCGCGCTAA 82  
Db 250 CTCCTGCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACGCGCTAA 191  
QY 83 TTTTGTATCTTTTAGTAGAGACGGCTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142  
Db 190 TTTTGTAT-TTTTAGTAGAGACGGGTTCTCCATGTTGATCAGGCTGGTCTTGAAGCTGC 132  
QY 143 AAACCTCAGGTGATCCCGCGCTCGCGCTCCCAAAGTGTAGATTACAGGCGTGAGCC 202  
Db 131 CAACCTCAGGTGATCCCGCGCTCGCGCTCCCAAAGTGTAGATTACAGGCTGAGCC 72  
QY 203 ACCGCGCTCAGCTGGGAACACCTTTTCTTACATCTTCAAGTG 245  
Db 71 ACCACGCCCGCGGCTCTTTATTTATGAAGCTTTGTGTG 29

## RESULT 9

US-09-621-976-14426/c  
; Sequence 14426, Application US/09621976

; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 14426  
; LENGTH: 373  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-14426

Query Match 34.4%; Score 153.4; DB 4; Length 373;  
Best Local Similarity 87.0%; Pred. No. 1.1e-30;  
Matches 180; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

QY 23 CTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACGCGCTAA 82  
Db 288 CTCCTGCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACGCGCTAA 229  
QY 83 TTTTGTATCTTTTAGTAGAGACGGCTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142  
Db 228 TTTTGTAT-TTTTAGTAGAGATGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 170  
QY 143 AAACCTCAGGTGATCCCGCGCTCGCGCTCCCAAAGTGTAGATTACAGGCGTGAGCC 202  
Db 169 CAACCTCAGGTGATCCCGCTCGCGCTCCCAAAGTGTAGATTACAGGCGTGAGCC 110  
QY 203 ACCGCGCTCAGCTGGGAACACCTTTT 229  
Db 109 ACCGCGCCAGCCATGTTGCAGCTCTT 83

## RESULT 10

PCT-US93-06251-80/c  
; Sequence 80, Application PC/TUS9306251  
; GENERAL INFORMATION:  
; APPLICANT: Wickstrom, Eric and Rife, Jason P.  
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: SCULLY, SCOTT, MURPHY & PRESSER  
; STREET: 400 Garden City Plaza  
; CITY: Garden City  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 11530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/06251  
; FILING DATE: 19930630  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Digiglio, Frank S.  
; REGISTRATION NUMBER: 31,346  
; REFERENCE/DOCKET NUMBER: 8586  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 516-742-4343  
; TELEFAX: 516-742-4366  
; TELEX: 230 901 SANS UR  
; INFORMATION FOR SEQ ID NO: 80:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 841 base pairs

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-80

Query Match          34.4%; Score 153.4; DB 5; Length 841;
Best Local Similarity 81.8%; Pred. No. 1.5e-30;
Matches 189; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

QY 23 CTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACCGCGGTAA 82
Db 250 CTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACCGGTAA 191
QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142
Db 190 TTTTGTAT-TTTTAGTAAAGATGGGTTTCCACATGTTGGTCAGGCTGGTCTCGAACTCC 132
QY 143 AAACCTCAGGTGATCGCGCGCTCGGCCTCCCAAAGTGTAGGATTACAGGCGGTGAGCC 202
Db 131 CGACCTCAGGTGATCTGCGCGCTTGGCTCCCAAAGTGTGGGATTACAGGTTGAGGCC 72
QY 203 ACCGCGCTCAGCTGGGAACACCTTTTCTTACATCTTCAAGTGTCTAGAAAT 253
Db 71 ACCGCACCCGCGCTGCTGCTGTTTTTGTGATCTTACAGGTCAAACTAAAT 21

RESULT 11
PCT-US93-06251-81/c
; Sequence 81, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Wickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patcin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06251
; FILING DATE: 19930630
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 81:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 841 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-81

Query Match          34.4%; Score 153.4; DB 5; Length 841;
Best Local Similarity 81.8%; Pred. No. 1.5e-30;
Matches 189; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACCGCGGTAA 82
```

```
Db 250 CTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACCGGTAA 191
QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142
Db 190 TTTTGTAT-TTTTAGTAAAGATGGGTTTCCACATGTTGGTCAGGCTGGTCTCGAACTCC 132
QY 143 AAACCTCAGGTGATCGCGCGCTCGGCCTCCCAAAGTGTAGGATTACAGGCGGTGAGCC 202
Db 131 CGACCTCAGGTGATCTGCGCGCTTGGCTCCCAAAGTGTGGGATTACAGGTTGAGGCC 72
QY 203 ACCGCGCTCAGCTGGGAACACCTTTTCTTACATCTTCAAGTGTCTAGAAAT 253
Db 71 ACCGCACCCGCGCTGCTGCTGTTTTTGTGATCTTACAGGTCAAACTAAAT 21

RESULT 12
US-09-544-398B-11
; Sequence 11, Application US/09544398B
; Patent No. 6770461
; GENERAL INFORMATION:
; APPLICANT: Carulli, John P.
; APPLICANT: Little, Randall D.
; APPLICANT: Recker, Robert R.
; APPLICANT: Johnson, Mark L.
; TITLE OF INVENTION: High bone mass gene of 11q13.3
; FILE REFERENCE: 032796-013
; CURRENT APPLICATION NUMBER: US/09/544,398B
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 09/229,319
; PRIOR FILING DATE: 1999-01-13
; PRIOR APPLICATION NUMBER: US 60/071,449
; PRIOR FILING DATE: 1998-01-13
; PRIOR APPLICATION NUMBER: US 60/105,511
; PRIOR FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 641
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 66933
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-544-398B-11

Query Match          34.4%; Score 153.4; DB 4; Length 66933;
Best Local Similarity 82.9%; Pred. No. 7.9e-30;
Matches 175; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACCGCGGTAA 82
Db 52539 CTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCAACCGCGGTAA 52598
QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142
Db 52599 TTTTGTATTTTGTAGTAGACAGAGGTTTCCACATGTTGGCCAGGCTGGTCTCGAACTCC 52658
QY 143 AAACCTCAGGTGATCGCGCGCTCGGCCTCCCAAAGTGTAGGATTACAGGCGGTGAGCC 202
Db 52659 TGACCTCAGGTGATCGCGCGCTCAGCCTCCCAAAGTGTGGGATTACAGGATGAGGCC 52718
QY 203 ACCGCGCTCAGCTGGGAACACCTTTTCTTCTTA 233
Db 52719 ACCGCGCGCGCTGAGTTTCTTTTATGA 52749

RESULT 13
US-09-543-771-11
; Sequence 11, Application US/09543771
; Patent No. 6780609
; GENERAL INFORMATION:
; APPLICANT: John P. Carulli et al.
; TITLE OF INVENTION: THE HIGH BONE MASS GENE OF 11q13.3
; FILE REFERENCE: 032796-013
; CURRENT APPLICATION NUMBER: US/09/543,771
```

```
; CURRENT FILING DATE: 2000-04-05
; EARLIER APPLICATION NUMBER: US 09/229,319
; EARLIER FILING DATE: 1999-01-13
; EARLIER APPLICATION NUMBER: US 60/071,449
; EARLIER FILING DATE: 1998-01-13
; EARLIER APPLICATION NUMBER: US 60/105,511
; EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 62
; SEQ ID NO 11
; LENGTH: 66933
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-543-771-11

Query Match      34.4%; Score 153.4; DB 4; Length 66933;
Best Local Similarity 82.9%; Pred. No. 7.9e-30;
Matches 175; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCCCGCTAA 82
Db 52539 CTCTGCTCAGCCTCCCGAGTAGCTGGGATGACAGGTGCGCACCACTGCTGGCTAA 52598

QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTTGCTCAGGCTGGTCTCGAACTTC 142
Db 52599 TTTTGTATTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCC 52658

QY 143 AAACCTCAGGTGATCGCCCGCTCGGCTCCCAAAGTCTAGGATTACAGGCGTGAGCC 202
Db 52659 TGACCTCAGGTGATCGCCCGCTCGGCTCCCAAAGTCTGGGATTACAGGCGTGAGCC 52718

QY 203 ACCGGCTCAGCCTGGGAACACCTTTTCTTA 233
Db 52719 ACCGGCCCGGCTGAGTTTCTTTTATGA 52749

RESULT 14
US-09-544-398B-9
; Sequence 9, Application US/09544398B
; Patent No. 6770461
; GENERAL INFORMATION:
; APPLICANT: Carulli, John P.
; APPLICANT: Little, Randall D.
; APPLICANT: Recker, Robert R.
; APPLICANT: Johnson, Mark L.
; TITLE OF INVENTION: High bone mass gene of 11q13.3
; FILE REFERENCE: 032796-013
; CURRENT APPLICATION NUMBER: US/09/544,398B
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 09/229,319
; PRIOR FILING DATE: 1999-01-13
; PRIOR APPLICATION NUMBER: US 60/071,449
; PRIOR FILING DATE: 1998-01-13
; PRIOR APPLICATION NUMBER: US 60/105,511
; PRIOR FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 641
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 72049
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (8356), (8385), (38585)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-544-398B-9
```

```
Query Match      34.4%; Score 153.4; DB 4; Length 72049;
Best Local Similarity 82.9%; Pred. No. 8.1e-30;
Matches 175; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCCCGCTAA 82
Db 56113 CTCTGCTCAGCCTCCCGAGTAGCTGGGATGACAGGTGCGCACCACTGCTGGCTAA 56172
```

```
QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTTGCTCAGGCTGGTCTCGAACTTC 142
Db 56173 TTTTGTATTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCC 56232

QY 143 AAACCTCAGGTGATCGCCCGCTCGGCTCCCAAAGTCTAGGATTACAGGCGTGAGCC 202
Db 56233 TGACCTCAGGTGATCGCCCGCTCGGCTCCCAAAGTCTGGGATTACAGGCGTGAGCC 56292

QY 203 ACCGGCTCAGCCTGGGAACACCTTTTCTTA 233
Db 56293 ACCGGCCCGGCTGAGTTTCTTTTATGA 56323

RESULT 15
US-09-543-771-9
; Sequence 9, Application US/09543771
; Patent No. 6780609
; GENERAL INFORMATION:
; APPLICANT: John P. Carulli et al.
; TITLE OF INVENTION: THE HIGH BONE MASS GENE OF 11q13.3
; FILE REFERENCE: 032796-013
; CURRENT APPLICATION NUMBER: US/09/543,771
; CURRENT FILING DATE: 2000-04-05
; EARLIER APPLICATION NUMBER: US 09/229,319
; EARLIER FILING DATE: 1999-01-13
; EARLIER APPLICATION NUMBER: US 60/071,449
; EARLIER FILING DATE: 1998-01-13
; EARLIER APPLICATION NUMBER: US 60/105,511
; EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 62
; SEQ ID NO 9
; LENGTH: 72049
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (8356), (8385), (38585)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-543-771-9
```

```
Query Match      34.4%; Score 153.4; DB 4; Length 72049;
Best Local Similarity 82.9%; Pred. No. 8.1e-30;
Matches 175; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCCCGCTAA 82
Db 56113 CTCTGCTCAGCCTCCCGAGTAGCTGGGATGACAGGTGCGCACCACTGCTGGCTAA 56172

QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTTGCTCAGGCTGGTCTCGAACTTC 142
Db 56173 TTTTGTATTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCC 56232

QY 143 AAACCTCAGGTGATCGCCCGCTCGGCTCCCAAAGTCTAGGATTACAGGCGTGAGCC 202
Db 56233 TGACCTCAGGTGATCGCCCGCTCGGCTCCCAAAGTCTGGGATTACAGGCGTGAGCC 56292

QY 203 ACCGGCTCAGCCTGGGAACACCTTTTCTTA 233
Db 56293 ACCGGCCCGGCTGAGTTTCTTTTATGA 56323
```

Search completed: January 15, 2005, 18:23:05  
Job time : 78 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 15, 2005, 18:21:58 ; Search time 385 Seconds  
(without alignments)

6656.276 Million cell updates/sec

Title: US-10-009-579-5\_COPY\_3115\_3560

Perfect score: 446

Sequence: 1 gtaatggcagcatctctgct.....ctggaaggtctctgctgct 446

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09E\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 20: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	446	100.0	4282	13	US-10-009-579-5	Sequence 5, Appli	
2	169.6	38.0	4388	10	US-09-764-891-7647	Sequence 7647, App	
3	169.2	37.9	3252	16	US-10-108-260A-604	Sequence 604, Appl	
c	4	166	37.2	26874	15	US-10-004-113-34	Sequence 34, Appl
c	5	165.8	37.2	289	14	US-10-115-278-4	Sequence 4, Appli
c	6	165.8	37.2	289	18	US-10-762-366-4	Sequence 4, Appli
c	7	165.8	37.2	291	15	US-10-229-058B-15	Sequence 15, Appl
c	8	164.4	36.9	133300	18	US-10-331-053-70	Sequence 70, Appl
9	163.6	36.7	912	13	US-10-027-632-165142	Sequence 165142,	
10	163.6	36.7	912	13	US-10-027-632-165143	Sequence 165143,	
11	163.6	36.7	912	15	US-10-037-632-165142	Sequence 165142,	
12	163.6	36.7	912	15	US-10-037-632-165143	Sequence 165143,	

13	163.4	36.6	96597	16	US-10-052-482-226	Sequence 226, App
c	14	163	36.5	1136	13	Sequence 117277,
c	15	163	36.5	1136	13	US-10-027-633-117278
c	16	163	36.5	1136	15	Sequence 117277,
c	17	163	36.5	1136	15	US-10-027-632-117277
c	18	163	36.5	1136	15	US-10-027-632-117278
c	19	163	36.5	66025	18	Sequence 117278,
c	20	162.8	36.5	1980090	18	Sequence 6945, Ap
c	21	162.8	36.5	49745	13	Sequence 6815, Ap
c	22	162.2	36.4	8220	9	Sequence 1930, Ap
c	23	162.2	36.4	8220	15	Sequence 2, Appli
c	24	162.2	36.4	8220	15	Sequence 3, Appli
c	25	161	36.1	160361	16	Sequence 35, Appl
c	26	160.4	36.0	83517	17	Sequence 1, Appli
c	27	160.2	35.9	1018	14	Sequence 50, Appl
c	28	160.2	35.9	1250	10	Sequence 14015, A
c	29	160	35.9	3062	15	Sequence 21091, A
c	30	160	35.9	3062	15	Sequence 50, Appl
c	31	160	35.9	6534	10	Sequence 5454, Ap
c	32	159.6	35.8	2448	13	Sequence 298, App
c	33	159.6	35.8	2448	15	Sequence 262796,
c	34	159.6	35.8	3656	15	Sequence 965, App
c	35	159.6	35.8	6670	10	Sequence 5534, Ap
c	36	159.6	35.8	7713	10	Sequence 5533, Ap
c	37	159.6	35.8	53623	18	Sequence 44, Appl
c	38	159.6	35.8	129042	13	Sequence 1240, Ap
c	39	159.6	35.8	174448	13	Sequence 148, App
c	40	159.4	35.7	288	14	Sequence 2, Appli
c	41	159.4	35.7	288	18	Sequence 2, Appli
c	42	159.4	35.7	91760	13	Sequence 844, App
c	43	159.4	35.7	168821	13	Sequence 622, App
c	44	159.2	35.7	19300	17	Sequence 4, Appli
c	45	159.2	35.7	91352	16	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-10-009-579-5  
; Sequence 5, Application US/10009579  
; Publication No. US20020156041A1  
; GENERAL INFORMATION:  
; APPLICANT: Leij de, Lou F.M.H.  
; APPLICANT: Ruiters, Marcel H.J.  
; APPLICANT: McLaughlin, Pamela M.J.  
; APPLICANT: Harmsen, Martin C.  
; APPLICANT: Molen v.d., Henk  
; APPLICANT: Terpstra, Peter  
; APPLICANT: Dokter, Willem H.A.  
; TITLE OF INVENTION: Non-squamous epithelium-specific transcription  
; FILE REFERENCE: P52075US00  
; CURRENT APPLICATION NUMBER: US/10/009,579  
; CURRENT FILING DATE: 2002-03-26  
; PRIOR APPLICATION NUMBER: EP 00200728.4  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: PCT/NL01/00166  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 4282  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(4282)  
; OTHER INFORMATION: /note="EGP-2 promoter sequence from -3967 to +315"  
US-10-009-579-5

Query Match 100.0%; Score 446; DB 13; Length 4282;  
Best Local Similarity 100.0%; Pred. No. 6.1e-98;  
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTAATGGCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGATTACAGGC 60  
Db 3115 GTAATGGCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGATTACAGGC 3174  
QY 61 ATGGCCACACACCGCCGGCTAAATTTGTATCTTTTAGTAGAGAGCGGTTCTCCCATGTT 120  
Db 3175 ATGGCCACACACCGCCGGCTAAATTTGTATCTTTTAGTAGAGAGCGGTTCTCCCATGTT 3234  
QY 121 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCCGCCCGCTCGGCTCCCAAAGT 180  
Db 3235 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCCGCCCGCTCGGCTCCCAAAGT 3294  
QY 181 GCTAGGATTACAGCGCTGAGCCACCGCGCTCAGCCTGGGAAACACCTTTCTTACATCTTC 240  
Db 3295 GCTAGGATTACAGCGCTGAGCCACCGCGCTCAGCCTGGGAAACACCTTTCTTACATCTTC 3354  
QY 241 AAGTGCTAGAAATGCTTTATGAAAACGAAAAAGAAATTTAAGAGTAATTTATAAGAAAC 300  
Db 3355 AAGTGCTAGAAATGCTTTATGAAAACGAAAAAGAAATTTAAGAGTAATTTATAAGAAAC 3414  
QY 301 ACTCAATTTCTCCCAAGAGAGCAAGATTTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 360  
Db 3415 ACTCAATTTCTCCCAAGAGAGCAAGATTTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 3474  
QY 361 TTTCTAATTTCAAGAGATATAATTAATTTGCCAGGTAAAGCTCAAGGTCTTTTTTAT 420  
Db 3475 TTTCTAATTTCAAGAGATATAATTAATTTGCCAGGTAAAGCTCAAGGTCTTTTTTAT 3534  
QY 421 AGTGTTCTGGAAGGTTCTCTGCGCTGT 446  
Db 3535 AGTGTTCTGGAAGGTTCTCTGCGCTGT 3560

## RESULT 2

US-09-764-891-7647  
; Sequence 7647, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; PRIOR FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7647  
; LENGTH: 4388  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-7647

Query Match 38.0%; Score 169.6; DB 10; Length 4388;  
Best Local Similarity 86.6%; Pred. No. 8e-31;  
Matches 187; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 GTAATGGCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGATTACAGGC 60  
Db 1079 GCAATGGCAGCATCTCTGCTCGTGCATCCCTCGCCTCCCGAGTAGCTGGATTACAGGC 1138  
QY 61 ATGGCCACACACCGCCGGCTAAATTTGTATCTTTTAGTAGAGAGCGGTTCTCCCATGTT 120  
Db 1139 GCGCGCCACACAGTGGCGGCTAAATTTGTATCTTTTAGTAGAGAGCGGTTTACCAATGTT 1198  
QY 121 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCCGCCCGCTCGGCTCCCAAAGT 180  
Db 1199 GGCCAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCCGCCCGCTCGGCTCCCAAAGT 1258  
QY 181 GCTAGGATTACAGCGCTGAGCCACCGCGCTCAGCCT 216  
Db 1259 GCTGGAATTACAGGCTTGAGTCACTCGCCCGCGCT 1294

RESULT 3  
US-10-108-260A-604  
; Sequence 604, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 604  
; LENGTH: 3252  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-108-260A-604

Query Match 37.9%; Score 169.2; DB 16; Length 3252;  
Best Local Similarity 81.9%; Pred. No. 8.8e-31;  
Matches 195; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1 GTAATGGCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGATTACAGGC 60  
Db 2742 GCAATGGCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGATTACAGGC 2801  
QY 61 ATGCGCCACACCGCCGGCTAAATTTGTATCTTTTAGTAGAGAGCGGTTCTCCCATGTT 120  
Db 2802 ACGTACCACACCGCTGGCTAATTTTGTATTTATTAGTAGAGAGCGGTTTCAACCATGTT 2861  
QY 121 GGTGAGGCTGGTCTCGAAGCTTCAAACTCAGGTGATCCGCCCGCTCGGCTCCCAAAGT 180  
Db 2862 GCGCCAGGCTGGTCTCAAACTCCTGACCTCAGTATCCACCTCCCTCAGCCTCCCAAAGT 2921  
QY 181 GCTAGGATTACAGGCTGAGCCACCGCTCAGCCTCAGCCTGGGAACACCTTTTCTTACATCT 238  
Db 2922 ACTGGATTACAGGCTGAGCCACTGCGCCCACTGCGCCCACTGAGAGCAGCTTTTAATACATTT 2979

## RESULT 4

US-10-004-113-34/c  
; Sequence 34, Application US/10004113  
; Publication No. US20030194702A1  
; GENERAL INFORMATION:  
; APPLICANT: Engelhard, Eric  
; APPLICANT: Morris, David  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: A-70970/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/10/004,113  
; PRIOR FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 34  
; LENGTH: 26874  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (25082)..(25101)  
; OTHER INFORMATION: "n" at positions 25082 through 25101 can be any base.  
US-10-004-113-34

Query Match 37.2%; Score 166; DB 15; Length 26874;  
Best Local Similarity 81.1%; Pred. No. 1.3e-29;  
Matches 193; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1 GTAATGGCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGATTACAGGC 60

Db 22835 GCAGTGGCATGATCTCAGCTCACTGCAACCTCAGCCCTCTCGAGTAGCTGGGATTACAGGC 22776  
QY 61 ATCGGCCACACGCGCGGCTAAATTTTCTATCTTTTAGTAGAGAGCGGTTCTCCCATGTT 120  
Db 22775 ACGTACACACGCGCTGGCTAAATTTTGTATTTTAGTAGAGACAGGTTTACCATGTT 22716  
QY 121 GGTGAGCTGGTCTCGAACTTCAAACTCAGGTGATCCGCCCGCTCGGCTCCCAAAGT 180  
Db 22715 GGCAGGCTGGTCTCAAACTCCTGACCTCAGTGATCCACCTCCTCAGCCTCCCAAAGT 22656  
QY 181 GCTAGATTACAGCGTAGCCACCGCGCTCAGCTGGGAAACCTTTTCTTACATCT 238  
Db 22655 ACTGGGATTACAAGCGTAGCCACTGCGCCCAACTGAGAAGCAGCTTTTAAATACATTT 22598

RESULT 5  
US-10-115-278-4/c  
; Sequence 4, Application US/10115278  
; Publication No. US2003082644A1  
; GENERAL INFORMATION:  
; APPLICANT: Schuetz, Ekkehard  
; APPLICANT: Urnovitz, Howard B. Inc.  
; APPLICANT: Chronix Biomedical, Inc.  
; TITLE OF INVENTION: Diagnostic Detection of Nucleic Acids  
; FILE REFERENCE: 018651-000320US  
; CURRENT APPLICATION NUMBER: US/10/115,278  
; CURRENT FILING DATE: 2002-07-09  
; PRIOR APPLICATION NUMBER: US 60/280,523  
; PRIOR FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 289  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:AluSp consensus  
US-10-115-278-4

Query Match 37.2%; Score 165.8; DB 14; Length 289;  
Best Local Similarity 91.2%; Pred. No. 2.1e-30;  
Matches 176; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCGGGCTAA 82  
Db 193 CTCCTGCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCGGGCTAA 134  
QY 83 TTTTGTATCTTTTAGTAGAGACGCGGTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142  
Db 133 TTTTGTATCTTTTAGTAGAGACGCGGTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 74  
QY 143 AAACCTCAGGTGATCCCGCGCTCGGCCCTCCCAAAGTGTAGGATTACAGGCGGTGAGCC 202  
Db 73 CGACCTCAGGTGATCCCGCGCTCGGCCCTCCCAAAGTGTGGGATTACAGGCGGTGAGCC 14  
QY 203 ACCGCGCTCAGCC 215  
Db 13 ACCGCGCGCGGCC 1

RESULT 6  
US-10-762-966-4/c  
; Sequence 4, Application US/10762966  
; Publication No. US20040241712A1  
; GENERAL INFORMATION:  
; APPLICANT: Schuetz, Ekkehard  
; APPLICANT: Urnovitz, Howard B. Inc.  
; APPLICANT: Chronix Biomedical, Inc.  
; TITLE OF INVENTION: Diagnostic Detection of Nucleic Acids  
; FILE REFERENCE: 018651-000320US  
; CURRENT APPLICATION NUMBER: US/10/762,966  
; CURRENT FILING DATE: 2004-01-21

; PRIOR APPLICATION NUMBER: US/10/115,278  
; PRIOR FILING DATE: 2002-07-09  
; PRIOR APPLICATION NUMBER: US 60/280,523  
; PRIOR FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 289  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:AluSp consensus  
US-10-762-966-4

Query Match 37.2%; Score 165.8; DB 18; Length 289;  
Best Local Similarity 91.2%; Pred. No. 2.1e-30;  
Matches 176; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCGGGCTAA 82  
Db 193 CTCCTGCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCGGGCTAA 134  
QY 83 TTTTGTATCTTTTAGTAGAGACGCGGTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142  
Db 133 TTTTGTATCTTTTAGTAGAGACGCGGTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 74  
QY 143 AAACCTCAGGTGATCCCGCGCTCGGCCCTCCCAAAGTGTAGGATTACAGGCGGTGAGCC 202  
Db 73 CGACCTCAGGTGATCCCGCGCTCGGCCCTCCCAAAGTGTGGGATTACAGGCGGTGAGCC 14  
QY 203 ACCGCGCTCAGCC 215  
Db 13 ACCGCGCGCGGCC 1

RESULT 7  
US-10-229-058B-15/c  
; Sequence 15, Application US/10229058B  
; Publication No. US20030194718A1  
; GENERAL INFORMATION:  
; APPLICANT: Hiroyuki TOMITA, Toshiro SAITO, Masatoshi NARAHARA and  
; APPLICANT: Hiroyasu KATO  
; TITLE OF INVENTION: PROBING SEQUENCE DETERMINATION SYSTEM FOR DNA  
; FILE REFERENCE: PH-1629  
; CURRENT APPLICATION NUMBER: US/10/229,058B  
; CURRENT FILING DATE: 2002-08-28  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 291  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-229-058B-15

Query Match 37.2%; Score 165.8; DB 15; Length 291;  
Best Local Similarity 91.2%; Pred. No. 2.1e-30;  
Matches 176; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCGGGCTAA 82  
Db 193 CTCCTGCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCGCGGGCTAA 134  
QY 83 TTTTGTATCTTTTAGTAGAGACGCGGTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142  
Db 133 TTTTGTATCTTTTAGTAGAGACGCGGTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 74  
QY 143 AAACCTCAGGTGATCCCGCGCTCGGCCCTCCCAAAGTGTAGGATTACAGGCGGTGAGCC 202  
Db 73 CGACCTCAGGTGATCCCGCGCTCGGCCCTCCCAAAGTGTGGGATTACAGGCGGTGAGCC 14  
QY 203 ACCGCGCTCAGCC 215

```
Db      13 ACCGCGCCGGCC 1
||||| | | | |
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165142
; LENGTH: 912
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-165142

Query Match      36.7%; Score 163.6; DB 13; Length 912;
Best Local Similarity 82.0%; Pred. No. 1.2e-29;
Matches 187; Conservative 1; Mismatches 40; Indels 0; Gaps 0;

QY      5 TGGCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGC 64
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      39 TGCCAGGTTCAAGCACTTCTCTGCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGT 98
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      65 GCCACCACGCCCGGCTAAATTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTTGGTC 124
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      99 GCCACCACGCCCTGGCTTAATTTTGTATTTTGTAGTAGACACGGGGTTTCTCCATGTTGGTC 158
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      125 AGGCTGGTCTCGAACTTCAACCTCAGGTGATCGCGCCGCTCGGCTCCCAAAGTGCTA 184
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      159 AGGCTAGTCTCGAACTTCCCACTCAGGTGATCGCGCCGCTCAGCCTCCCAAAGTGCTG 218
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      185 GGATTACAGCGCTGAGCCACCGCGCTCAGCCTGGGAAACACCTTTTCTT 232
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      219 GGATTACAGCGCTGAACCACTCGGCCAGCGCTGGTTGGACCAATATTTT 266
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
US-10-027-632-165143
; Sequence 165143, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165143
; LENGTH: 912
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-165143

Query Match      36.7%; Score 163.6; DB 13; Length 912;
Best Local Similarity 82.0%; Pred. No. 1.2e-29;
Matches 187; Conservative 1; Mismatches 40; Indels 0; Gaps 0;

QY      5 TGGCAGCATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGC 64
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      39 TGCCAGGTTCAAGCACTTCTCTGCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGT 98
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      65 GCCACCACGCCCGGCTAAATTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTTGGTC 124
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      99 GCCACCACGCCCTGGCTTAATTTTGTATTTTGTAGTAGACACGGGGTTTCTCCATGTTGGTC 158
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-10-027-632-165142
; Sequence 165142, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
```



		184
Qy	125	AGCGTGGTCTCGAACTTCAAACCTCAGGTGATCGCCCGGCTCGGCCTCCCAAGTGCTA
		184
Db	159	AGCGTAGTCTCGAACTCCCAACCTCAGGTGATCGCCCGGCTCAGCCTCCCAAGTGCTG
		218
Qy	185	GGATTACAGCGGTGAGCCACCGCGCTCAGCGCTGGGAACAGCTTTTCTT
		232
Db	219	GGATTACAGGCGTGAACCACTCGGCCAGCGCTGGTGGACATATTT
		266

```

RESULT 11
US-10-027-632-165142
; Sequence 165142, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165142
; LENGTH: 912
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-165142

```

```

RESULT 12
US-10-027-632-165143
; Sequence 165143, Application US/10027632
; Publication NO. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129

```

```
; OTHER INFORMATION: "n" at positions 13303 to 21086 can be any base
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (30929)..(31963)
; OTHER INFORMATION: "n" at positions 30929 to 31963 can be any base
US-10-052-482-226

Query Match          36.6%; Score 163.4; DB 16; Length 96597;
Best Local Similarity 87.6%; Pred. No. 9.5e-29;
Matches 190; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

QY 1 GTAATGCGACGATCTCTGCTCACTGCAACCTCAGCCTCCCGAGTAGCTGGATTACAGGC 60
Db 76411 GCAGTGGCGTGAATCTCGCTCAGCTCACTGCAACCTCCAGTAGCTGGATTACAGGC 76470

QY 61 ATGGCCACACAGCCCGCGCTAAATTTTCTATCTTTTAGTAGAGCGGCTTCCTCCATGTT 120
Db 76471 ATGCACACACAGCCCGCGCTAAATTTTCTA-ATTTAGTAGAGATGGGTTTTCGCATGTT 76529

QY 121 GGTGAGGCTGGTCTCGAAGCTTCAACCTCAGGTGATCGCGCGCTCGGCTCCCAAGT 180
Db 76530 GGCCAGCAGGCTCTGGAACCTCCCAACCTCAGGTGATCGCGCGCTCGGCTCCCAAGT 76589

QY 181 GCTAGGATTACAGCGCTGAGCCAGCCGCGCTCAGCCTG 217
Db 76590 GCTGGGATTACAGCGATGAGCCAGCCGCGCGCTG 76626

RESULT 14
US-10-027-632-117277/c
; Sequence 117277, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117277
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-117277

Query Match          36.5%; Score 163; DB 13; Length 1136;
Best Local Similarity 67.5%; Pred. No. 1.8e-29;
Matches 226; Conservative 2; Mismatches 107; Indels 0; Gaps 0;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGATTACAGGCATGCGCCACCGCCGCTAA 82
Db 418 CTCCTGCTCAGCCTCCTGAGTAGCTGGATTACAGCATGACCAACCGCCGCTAA 359

QY 83 TTTTGTATCTTTTAGTAGAGCGGCTTCCTCCATGTTGGTCTCAGGCTGCTCGAATTC 142
Db 358 TTTTGTATTTTGTAGTAGTAGGCTTCACCATGCTGGTCTGAGCTGCTTGAATCC 299

QY 143 AAACCTCAGGTGATCGCGCGCTCGGCTCCCAAGTCTAGGATTACAGGCGTAGCC 202
Db 298 CAACCTCAGGTGATCGCGCTCCCAACCTCCCAAGTCTAGGATTACAGGCGTAGCC 239

QY 203 ACCGGCTCAGCCTCGGGAACACCTTTTCTTACATCTTCAAGTCTAGAAATGCTTATGAA 262
Db 238 ACCGGCGCGGCTCACTACCCCTTTTCTATATTACAATATGAACAATTAATTATGAAA 179

QY 263 AACGAAAAAAGAAATTTAAGAGTAATTAATAAGAAACACTCATTTTCTCCCAAGAG 322
Db 178 AAATAATATGGCATGATGGGGGAGCAGAAATGGTAGCTCATCATGTTTGGAAAGTAGAAA 119

323 CCAAGATTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 357
118 CCAGCAGATAAAATATAGACATGTCTATCTATGTT 84
```

```
RESULT 15
US-10-027-632-117278/c
; Sequence 117278, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117278
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-117278

Query Match          36.5%; Score 163; DB 13; Length 1136;
Best Local Similarity 67.5%; Pred. No. 1.8e-29;
Matches 226; Conservative 2; Mismatches 107; Indels 0; Gaps 0;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGATTACAGGCATGCGCCACCGCCGCTAA 82
Db 418 CTCCTGCTCAGCCTCCTGAGTAGCTGGATTACAGCATGACCAACCGCCGCTAA 359

QY 83 TTTTGTATCTTTTAGTAGAGCGGCTTCCTCCATGTTGGTCTCAGGCTGCTCGAATTC 142
Db 358 TTTTGTATTTTGTAGTAGTAGGCTTCACCATGCTGGTCTGAGCTGCTTGAATCC 299

QY 143 AAACCTCAGGTGATCGCGCGCTCGGCTCCCAAGTCTAGGATTACAGGCGTAGCC 202
Db 298 CAACCTCAGGTGATCGCGCTCCCAACCTCCCAAGTCTAGGATTACAGGCGTAGCC 239

QY 203 ACCGGCTCAGCCTCGGGAACACCTTTTCTTACATCTTCAAGTCTAGAAATGCTTATGAA 262
Db 238 ACCGGCGCGGCTCACTACCCCTTTTCTATATTACAATATGAACAATTAATTATGAAA 179

QY 263 AACGAAAAAAGAAATTTAAGAGTAATTAATAAGAAACACTCATTTTCTCCCAAGAG 322
Db 178 AAATAATATGGCATGATGGGGGAGCAGAAATGGTAGCTCATCATGTTTGGAAAGTAGAAA 119
```



**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 15, 2005, 16:41:36 ; Search time 2254 Seconds  
(without alignments)  
7210.346 Million cell updates/sec

Title: US-10-009-579-5\_COPY\_3115\_3560

Perfect score: 446

Sequence: 1 gtaatggcagcatctctgct.....ctggaaggtctctgctgt 446

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_hic.\*

4: gb\_est3.\*

5: gb\_est4.\*

6: gb\_est5.\*

7: gb\_est6.\*

8: gb\_gss1.\*

9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	167.4	37.5	598	8 AQ416484	AQ416484 RPCI-11-1
C 2	161.8	36.3	482	1 AI148840	AI148840 QC65G11.x
C 3	161.4	36.2	660	5 BM997829	BM997829 UI-H-D10-
C 4	161	36.1	511	2 BE464585	BE464585 h85a05.x
C 5	160.6	36.0	624	8 AQ419825	AQ419825 RPCI-11-1
C 6	160.2	35.9	392	1 AI312614	AI312614 qp77c01.x
C 7	160.2	35.9	445	1 AI889177	AI889177 wm46g01.x
C 8	160.2	35.9	467	2 BF790866	BF790866 602250663
C 9	160.2	35.9	529	2 BF725436	BF725436 bx16a01.y
C 10	160.2	35.9	566	2 AW575808	AW575808 UI-HF-BNO
C 11	160.2	35.9	684	4 BM792342	BM792342 K-EST0072
C 12	160.2	35.9	740	4 BI087063	BI087063 602850725
C 13	160.2	35.9	906	4 BG259634	BG259634 602378774
C 14	160	35.9	714	7 CN426512	CN426512 170004706
C 15	159.8	35.8	1686	3 CR619941	CR619941 full-leng
C 16	159.8	35.8	1755	3 CR609780	CR609780 full-leng
C 17	159.2	35.7	477	2 BE929881	BE929881 RCS-GN004
C 18	159.2	35.7	486	2 BE049409	BE049409 xw85g12.x
C 19	159	35.7	418	1 AA496941	AA496941 ae32d09.s
C 20	159	35.7	499	5 BP871836	BP871836 BP871836
C 21	159	35.7	543	1 AL708895	AL708895 DKFZp686N
C 22	159	35.7	582	5 BP871862	BP871862 BP871862
C 23	159	35.7	608	5 BP871721	BP871721 BP871721
C 24	159	35.7	721	1 AV714079	AV714079 AV714079

25	159	35.7	1006	4 BM810643	BM810643 AGENCOURT
26	158.8	35.6	741	5 BX414857	BX414857 BX414857
C 27	158.6	35.6	321	7 F34164	F34164 HSPD28890 H
28	158.6	35.6	472	6 CA434300	CA434300 UI-H-DH0-
29	158.4	35.5	350	7 CR823193	CR823193 iJ25f02.y
C 30	158.4	35.5	393	1 AI472736	AI472736 ta13c04.x
C 31	158.4	35.5	398	2 BF475466	BF475466 nac30h01.
C 32	158.4	35.5	411	5 BQ101225	BQ101225 iJ25f02.y
C 33	158.4	35.5	421	2 AW440568	AW440568 xt15e04.x
C 34	158.4	35.5	422	2 BE892611	BE892611 601433420
C 35	158.4	35.5	438	1 AA678932	AA678932 ah08903.s
C 36	158.4	35.5	449	2 AW081610	AW081610 xc32b10.x
C 37	158.4	35.5	460	2 BE301610	BE301610 bb75e09.x
C 38	158.4	35.5	471	2 BE677244	BE677244 7d92d09.x
C 39	158.4	35.5	493	5 BU198009	BU198009 EA25 Subt
C 40	158.4	35.5	493	7 CN415426	CN415426 170004241
C 41	158.4	35.5	498	5 BM995211	BM995211 UI-H-ED0-
C 42	158.4	35.5	546	1 AI889995	AI889995 wm80c03.x
C 43	158.4	35.5	606	5 BQ778458	BQ778458 il131d07.x
C 44	158.4	35.5	619	5 BX505458	BX505458 DKFZp686E
C 45	158.4	35.5	656	6 CA420015	CA420015 UI-H-PH0-

ALIGNMENTS

RESULT 1  
AQ416484/c  
LOCUS  
DEFINITION RPCI-11-153H2.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-153H2, genomic survey sequence.  
ACCESSION AQ416484  
VERSION AQ416484.1 GI:4470608  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 598)  
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
JOURNAL Unpublished (1997)  
COMMENT Other GSSs: RPCI11-153H2.TJ  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeetigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet.cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1. .598  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7558537"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-153H2"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_11b="RPCI-11"  
/notes="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"

ORIGIN

Query Match 37.5%; Score 167.4; DB 8; Length 598;  
 Best Local Similarity 82.4%; Pred. No. 4.7e-19;  
 Matches 192; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1 GTAAATGCACGATCTCTGCTCACTGCAACCTCAGCCCTCCAGTAGCTGGGATTACAGGC 60  
 |||||  
 Db 343 GCAATGGCAATCTCTGCTCACTGCAACCTCTGCTCCCGAGTAGCTGGGATTATAGGC 284  
 |||||  
 QY 61 ATGCGCCACACGCGCCGGCTAAATTTTGTATCTTTTAGTAGAGACGGGTTCTCCCATGTT 120  
 |||||  
 Db 283 GTGCACCAACACGCGCCAGCTAAATTTTGTATTTTATTAGTAGAGACAAGTTTTCACCATGTT 224  
 |||||

QY 121 GGTCAAGCTGGTCTCGAACTTCAAACTCAGGTGATCCGCGCTCGGCTCCCAAAGT 180  
 |||||  
 Db 223 GGCAGAGCTGGTATCAAACTCCGTGACCTCAGGTGATCTGCCAGCTTGGCTCCCAAAGT 164  
 |||||

QY 181 GCTAGGATTACAGCGCTGAGCCAGCGCTCAGCTGGGAACACCTTTTCTTA 233  
 |||||

Db 163 GCTGGGATTACAGGTGTGAGCCACCATGCCAGCTGTGGCACACATTTCTGA 111  
 |||||

RESULT 2  
 A1148840  
 LOCUS  
 DEFINITION q65911.x1 Soares\_placenta 8to9weeks\_2Nbp8to9W Homo sapiens CDNA  
 clone IMAGE:1714532 3' similar to contains Alu repetitive element;;  
 mRNA sequence.

ACCESSION A1148840  
 VERSION A1148840.1 GI:3677309  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1027 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 422.

FEATURES  
 Location/Qualifiers  
 1..482  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1714532"  
 /dev\_stage="two placenta; one from 8 weeks and another  
 from 9 weeks post conception"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /clone\_lib="Soares\_placenta\_8to9weeks\_2Nbp8to9W"  
 /note="Organ: placenta; Vector: pT73D (Pharmacia) with a  
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 TGTACCAATCTGAAGTGGGCGCGCGATTTTTTTTTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library constructed by Bento Soares and  
 M.Fatima Bonaldo."

ORIGIN  
 Query Match 36.3%; Score 161.8; DB 1; Length 482;  
 Best Local Similarity 85.3%; Pred. No. 4.5e-18;  
 Matches 192; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 23 CTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCCACCACCGCGCTAA 82

Db 97 CTCCTGCCTCACCCCTCCCGAGTAGCTGGGATTACAGGCATGCCACCACCGCGCCAA 156  
 |||||  
 QY 83 TTTTGTATCTTTTAGTAGAGACGGGTTCTCTCATGTTGGTCAGGCTGTCTCGAACTTC 142  
 |||||  
 Db 157 TTTTGTAT-TTTTAGTAGAGACGGGTTCTCTCATGTTGGTCAGGCTGTCTCCAACCTC 215  
 |||||  
 QY 143 AAACCTCAGGTGATCCGCGCTCGGCTCCCAAGTCTAGGATTACAGCGGTGAGCC 202  
 |||||  
 Db 216 CAACCTCAGGTGATCCACCTGCTGGCTCCCAAAGTCTGGGATTACAGCGGTGTGCC 275  
 |||||  
 QY 203 ACCGCGCTCAGCTCGGAACACCTTTTCTTACATCTTCAAGTCT 247  
 |||||  
 Db 276 ACCACGCCACGCTTGGGTGTTTTTCTTTTCAGCTCTCCAGTACT 320  
 |||||

RESULT 3

BM997829

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BM997829 660 bp mRNA linear EST 17-JUN-2002  
 UI-H-D10-auw-g-20-0-UI.s1 NCI\_CGAP\_D10 Homo sapiens CDNA clone  
 IMAGE:5875243 3', mRNA sequence.

BM997829  
 EST.  
 BM997829.1 GI:19722730  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Jose Mercuende  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

The following repetitive elements were found in this CDNA  
 sequence: 11-303, >ALU (matched complement) 498-660, >ALU (matched  
 complement)  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

FEATURES  
 Location/Qualifiers  
 1..660  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5875243"  
 /tissue\_type="Lung Focal Fibrosis"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI\_CGAP\_D10"  
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: Eco RI; Site 2: Not I;  
 NCI\_CGAP D10 is a CDNA library containing the following  
 tissue(s): A pool of Lung Focal Fibrosis. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an Eco RI  
 adaptor, digested with Not I, and cloned directionally  
 into pT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 ATACCGGTC.

TAG\_TISSUE=Lung with fibrosis

TAG\_LIB=UI-H-D10

TAG\_SEQ=ATACCGGTC"

Matches		179;	Conservative	0;	Mismatches	30;	Indels	0;	Gaps	0;
QY	23	CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACCGCCGGCTAA 82								
Db	92	CTCCTGCCTCAGCGTCCCAAGTAGCTGGATTATAGGCATGTGCCACACCGCTGGCTAA 151								
QY	83	TTTGTGATCTTTTAGTAGAGAGCGGTTCTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142								
Db	152	TTTGTGATCTTTTAGCAGAGACTGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 211								
QY	143	AAACCTCAGGTGATCGCGCCCTCGGCTCCCAAAGTGTCCAAAGTGTAGGATTACAGCGCTGAGCC 202								
Db	212	CGACCTCAGGTGATCCATCTGCTCGGCTCCCAAAGTGTCCAAAGTGTAGGATTACAGCGCTGAGCC 271								
QY	203	ACCGCGCTCAGCTCGGGAACACCTTTTCT 231								
Db	272	ACCGTGCCTGGCCAGAGTACACTTTTAT 300								
RESULT 5										
LOCUS										
DEFINITION										
ACCESSION										
VERSION										
KEYWORDS										
SOURCE										
ORGANISM										
REFERENCE										
AUTHORS										
TITLE										
JOURNAL										
COMMENT										
FEATURES										
source										
ORIGIN										
Query Match										
Best Local Similarity										
Matches										
QY	23	CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACCGCCGGCTAA 82								
Db	111	CTCCTGCCTCAGCGTCCCAAGTAGCTGGATTATAGGCATGTGCCACACCGCTGGCTAA 170								
QY	83	TTTGTGATCTTTTAGTAGAGAGCGGTTCTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142								
Db	171	TTTGTGATCTTTTAGTAGAGAGCGGTTCTCGCATGTTGGTCAGGCTGGTCTCGAACTTC 230								
QY	143	AAACCTCAGGTGATCGCGCCCTCGGCTCCCAAAGTGTAGGATTACAGCGCTGAGCC 202								
Db	231	CAACCTCAGGTGATCGCGCCCTCGGCTCCCAAAGTGTGGGATTACAGCGCTGAGCC 290								
QY	203	ACCGCGCTCAGCTCGGGAACACCTTTTCTTACATCTTCAAGTGTAGGAAATGCTTATGAA 262								
Db	291	ACTCGGCCCAACCTTAATGGCTTTTGGTATTAAACACGAGACTCTGAATGTGCATCAA 350								
QY	263	AAGCAAAAGAAATTATTA 281								
Db	351	TAATCAATAATTTTATA 369								
RESULT 4										
LOCUS										
DEFINITION										
ACCESSION										
VERSION										
KEYWORDS										
SOURCE										
ORGANISM										
REFERENCE										
AUTHORS										
TITLE										
JOURNAL										
COMMENT										
FEATURES										
source										
ORIGIN										
Query Match										
Best Local Similarity										
Matches										
QY	23	CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACCGCCGGCTAA 82								
Db	92	CTCCTGCCTCAGCGTCCCAAGTAGCTGGATTATAGGCATGTGCCACACCGCTGGCTAA 151								
QY	83	TTTGTGATCTTTTAGTAGAGAGCGGTTCTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142								
Db	152	TTTGTGATCTTTTAGCAGAGACTGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 211								
QY	143	AAACCTCAGGTGATCGCGCCCTCGGCTCCCAAAGTGTCCAAAGTGTAGGATTACAGCGCTGAGCC 202								
Db	212	CGACCTCAGGTGATCCATCTGCTCGGCTCCCAAAGTGTCCAAAGTGTAGGATTACAGCGCTGAGCC 271								
QY	203	ACCGCGCTCAGCTCGGGAACACCTTTTCT 231								
Db	272	ACCGTGCCTGGCCAGAGTACACTTTTAT 300								
RESULT 5										
LOCUS										
DEFINITION										
ACCESSION										
VERSION										
KEYWORDS										
SOURCE										
ORGANISM										
REFERENCE										
AUTHORS										
TITLE										
JOURNAL										
COMMENT										
FEATURES										
source										
ORIGIN										
Query Match										
Best Local Similarity										
Matches										
QY	23	CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACCGCCGGCTAA 82								
Db	92	CTCCTGCCTCAGCGTCCCAAGTAGCTGGATTATAGGCATGTGCCACACCGCTGGCTAA 151								
QY	83	TTTGTGATCTTTTAGTAGAGAGCGGTTCTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142								
Db	152	TTTGTGATCTTTTAGCAGAGACTGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 211								
QY	143	AAACCTCAGGTGATCGCGCCCTCGGCTCCCAAAGTGTCCAAAGTGTAGGATTACAGCGCTGAGCC 202								
Db	212	CGACCTCAGGTGATCCATCTGCTCGGCTCCCAAAGTGTCCAAAGTGTAGGATTACAGCGCTGAGCC 271								
QY	203	ACCGCGCTCAGCTCGGGAACACCTTTTCT 231								
Db	272	ACCGTGCCTGGCCAGAGTACACTTTTAT 300								
RESULT 5										
LOCUS										
DEFINITION										
ACCESSION										
VERSION										
KEYWORDS										
SOURCE										
ORGANISM										
REFERENCE										
AUTHORS										
TITLE										
JOURNAL										
COMMENT										
FEATURES										
source										
ORIGIN										
Query Match										
Best Local Similarity										
Matches										
QY	23	CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACCGCCGGCTAA 82								
Db	92	CTCCTGCCTCAGCGTCCCAAGTAGCTGGATTATAGGCATGTGCCACACCGCTGGCTAA 151								
QY	83	TTTGTGATCTTTTAGTAGAGAGCGGTTCTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 142								
Db	152	TTTGTGATCTTTTAGCAGAGACTGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAACTTC 211								
QY	143	AAACCTCAGGTGATCGCGCCCTCGGCTCCCAAAGTGTCCAAAGTGTAGGATTACAGCGCTGAGCC 202								
Db	212	CGACCTCAGGTGATCCATCTGCTCGGCTCCCAAAGTGTCCAAAGTGTAGGATTACAGCGCTGAGCC 271								
QY	203	ACCGCGCTCAGCTCGGGAACACCTTTTCT 231								
Db	272	ACCGTGCCTGGCCAGAGTACACTTTTAT 300								
RESULT 5										
LOCUS										
DEFINITION										
ACCESSION										
VERSION										
KEYWORDS										
SOURCE										
ORGANISM										
REFERENCE										
AUTHORS										
TITLE										
JOURNAL										
COMMENT										
FEATURES										
source										
ORIGIN										
Query Match										
Best Local Similarity										
Matches										
QY	23	CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACCGCCGGCTAA 82								
Db	92	CTCCTGCCTCAGCGTCCCAAGTAGCTGGATTATAGGCATGTGCCACACCGCTGGCTAA 151								
QY	83	TTTGTGATCTTTTAGTAGAGAGCGGTTCTCTCCATGTTGGTCAGGCTGGTCTCG								

```

Db      205 CTCCTGCCTCAGGCTCCCAAGTAGCTGGGATTACAGCGTGCACACGACGCTCCGGCTAA 264
QY      83 TTTTGTATCTTTTAGTAGAGAGCGGCTTCCTCCATGTTGGTTCAGGCTGGTCTCGAACTTC 142
Db      265 TTTTGTATCTTTTAGTAGAGATGGAGTTTCGCCATGTTGGCCAGGCTAGTCTTGAATCC 324
QY      143 AAACCTCAGGTGATCCGCCCGCTCGGCTCCCAAAGTGTAGGATTACAGGCGTGAGCC 202
Db      325 TGACCTCAGGTGATCCACTGCTCGGCTCCCAAAGTGTGGGATTACAGGCGTGAGCC 384
QY      203 ACCGGCTCAGCTGGGGAACCTTTTCTTACATCTTCAAGTGTAGAAATGCTTATGAA 262
Db      385 ACCAGCCAGCCAGGAACTCCTTCTTAATGATTCTTCCCTCAATCTCCTGTTGA 444
QY      263 AACGAAAAAGAAAT 277
Db      445 AAGGAGAAAAAGTTT 459

RESULT 6
LOCUS   AI1312614
DEFINITION gp77c01.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:1929024 3' similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION AI1312614
VERSION   AI1312614.1 GI:4018219
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 392)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 815 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 378.
FEATURES             Location/Qualifiers
     source           1..392
     organism="Homo sapiens"
     mol_type="mRNA"
     db_xref="taxon:9606"
     clone="IMAGE:1929024"
     dev_stage="19 weeks"
     lab_hosts="DH10B (ampicillin resistant)"
     clone_lib="Soares_fetal_lung_NbHL19W"
     notes="Organ: lung; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NbHL19W."

ORIGIN
Query Match      35.9%; Score 160.2; DB 1; Length 392;
Best Local Similarity 84.9%; Pred. No. 8.ee-18;
Matches 191; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY      23 CTCGAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCACACCGCCGGCTAA 82

```

```

Db      94 CTCCTGCCTCACCCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACGCGCGCCAA 153
QY      83 TTTTGTATCTTTTAGTAGAGAGCGGCTTCCTCCATGTTGGTTCAGGCTGGTCTCGAACTTC 142
Db      154 TTTTGTAT-TTTTAGTAGAGACGGGGTTTCTCCATGTTGGTCAAGCTGGTCTCCAATCC 212
QY      143 AAACCTCAGGTGATCCGCCCGCTCGGCTCCCAAAGTGTAGGATTACAGGCGTGAGCC 202
Db      213 CAACCTCAGGTGATCCACTGCTCGGCTCCCAAAGTGTGGGATTACAGGCGTGAGCC 272
QY      203 ACCGGCTCAGCTGGGAACACCTTTTCTTACATCTTCAAGTGCT 247
Db      273 ACCAGCCAGCGCTTGGGTGTTTTTCTTACGCTCTCCAGTACT 317

RESULT 7
LOCUS   AI889177
DEFINITION wm46g01.x1 NCI_CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2439024 3' similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION AI889177
VERSION   AI889177.1 GI:5594341
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 445)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1983 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source           1..445
     organism="Homo sapiens"
     mol_type="mRNA"
     db_xref="taxon:9606"
     clone="IMAGE:2439024"
     tissue_type="serous papillary carcinoma, high grade, 2 pooled tumors"
     lab_host="DH10B"
     clone_lib="NCI_CGAP_Ut4"
     notes="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.48 kb. Life Technologies catalog #: 11542-016"

ORIGIN
Query Match      35.9%; Score 160.2; DB 1; Length 445;
Best Local Similarity 84.9%; Pred. No. 8.5e-18;
Matches 191; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY      23 CTGAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCACACCGCCGGCTAA 82
Db      98 CTCCTGCCTCACCCTCCCGAGTAGCTGGGATTACAGGCATGCGCACACCGCCGGCCAA 157
QY      83 TTTTGTATCTTTTAGTAGAGAGCGGCTTCCTCCATGTTGGTTCAGGCTGGTCTCGAACTTC 142
Db      158 TTTTGTAT-TTTTAGTAGAGACGGGGTTTCTCCATGTTGGTCAAGCTGGTCTCCAATCC 216

```



Qy 143 AAACCTCAGGTGATCGCGCCCTCGGCTCCCAAAGTGCTAGGATTACAGCGGTGAGCC 202  
 |||||  
 Db 217 CAACCTCAGGTGATCCACCTCGGCTCGGCTCCCAAAGTGCTGGGATTACAGCGGTGAGCC 276  
 |||||  
 Qy 203 ACCGCGTCTAGCCTGGGAACACCTTTTCTTACATCTTCAAGTGCT 247  
 |||||  
 Db 277 ACCAGGCCAGCCTTGGGTGTTTCTTCTTACGCTCCTCCAGTACT 321  
 |||||

RESULT 8  
 BF790866/c  
 LOCUS 602250663P1 NIH\_MGC\_81 Homo sapiens cDNA clone IMAGE:4338231 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BF790866  
 VERSION BF790866.1 GI:12095920  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 467)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LICM1213 row: m column: 16  
 High quality sequence stop: 439.

FEATURES  
 source  
 1..467  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4338231"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIH\_MGC\_81"  
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB  
 (Clontech); Site\_1: SfiI (ggccgctcgccg); Site\_2: SfiI  
 (ggccattagggc); 5' and 3' adaptors were used in cloning  
 as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3',  
 and 3' adaptor sequence:  
 5'-ATTCTAGAGCGGCGGCGGCGGACATG-dT(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size  
 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained  
 inserts by PCR. This library was enriched for full-length  
 clones and was constructed by Clontech Laboratories (Palo  
 Alto, CA)."

ORIGIN  
 Query Match 35.9%; Score 160.2; DB 2; Length 467;  
 Best Local Similarity 84.9%; Pred. No. 8.4e-18;  
 Matches 191; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

Qy 23 CTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACCGCGGCTAA 82  
 |||||  
 Db 342 CTCCTGCTACCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACCGCGGCAA 283  
 |||||  
 Qy 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCTCCATGTTGGTTCAGGCTGGTCTCGAACTTC 142  
 |||||  
 Db 282 TTTTGTAT-TTTTAGTAGAGACGGGTTTCTCCATGTTGGTTCAGGCTGGTCTCGAACTTC 224  
 |||||  
 Qy 143 AAACCTCAGGTGATCGGCGCCCTCGGCTCCCAAAGTGCTAGGATTACAGCGGTGAGCC 202  
 |||||  
 Db 223 CAACCTCAGGTGATCCACCTCGGCTCGGCTCCCAAAGTGCTGGGATTACAGCGGTGAGCC 164  
 |||||

Qy 203 ACCGCGTCTAGCCTGGGAACACCTTTTCTTACATCTTCAAGTGCT 247  
 |||||  
 Db 163 ACCAGGCCAGCCTTGGGTGTTTCTTCTTACGCTCCTCCAGTACT 119  
 |||||

RESULT 9  
 BF725436/c  
 LOCUS BF725436  
 DEFINITION bxl6a01.y1 Human Iris cDNA (Un-normalized, unamplified): BX Homo  
 sapiens cDNA clone bxl6a01 5', mRNA sequence.  
 ACCESSION BF725436  
 VERSION BF725436.1 GI:12041355  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 529)  
 AUTHORS Wistow, G.J., Bernstein, S., Behal, A. and Smith, D.  
 TITLE NIBANK: EST analysis and bioinformatics for ocular genomics  
 JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press  
 COMMENT Contact: Wistow G

Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: [graeme@helix.nih.gov](mailto:graeme@helix.nih.gov)  
 Plate: 16 row: a column: 01  
 Seq primer: M13RP1 reverse primer (ABI).

FEATURES  
 source  
 1..529  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="bxl6a01"  
 /tissue\_type="Iris"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="Human Iris cDNA (Un-normalized, unamplified):  
 BX"

/note="Organ: Eye; Vector: pCMVSPORT6; Post-mortem iris  
 tissue was pooled from 10 individuals ranging in age from  
 4-80 years and RNA was extracted. From this pooled sample  
 an aliquot of 60ug of total RNA yielded 2.17ug of mRNA. A  
 directionally cloned cDNA library in the pCMVSPORT6 vector  
 was constructed at Life Technologies, essentially  
 following the protocols of the SuperScript Plasmid System  
 full details of which are contained in the manufacturer's  
 instruction manual (<http://www.lifetech.com/>). First  
 strand synthesis was carried out using a Not I  
 primer-adaptor  
 [5'-pGACTAGTTCATAGTCGAGCGCGCC(T)15-3']. Not I/blunt  
 end inserts were cloned into the Not I/EcoR V sites in the  
 vector. EST analysis was performed on the unamplified  
 library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN  
 Query Match 35.9%; Score 160.2; DB 2; Length 529;  
 Best Local Similarity 84.9%; Pred. No. 8.2e-18;  
 Matches 191; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

Qy 23 CTGCAACCTCAGCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACCGCGGCTAA 82  
 |||||  
 Db 331 CTCCTGCTCACCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACACCGCGGCAA 272  
 |||||  
 Qy 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCTCATGTTGGTTCAGGCTGGTCTCGAACTTC 142  
 |||||  
 Db 271 TTTTGTAT-TTTTAGTAGAGACGGGTTTCTCCATGTTGGTTCAGGCTGGTCTCGAACTTC 213  
 |||||  
 Qy 143 AAACCTCAGGTGATCGGCGCCCTCGGCTCCCAAAGTGCTAGGATTACAGCGGTGAGCC 202  
 |||||  
 Db 212 CAACCTCAGGTGATCCACCTCGGCTCGGCTCCCAAAGTGCTGGGATTACAGCGGTGAGCC 153  
 |||||

```

QY 203 ACCGGCTCAGCTGGGACACCTTTTCTTACATCTTCAAGTGCT 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 287 ACCACGCCAGCGCTGGGTGTTTCTTTCAGCTCTCCAGTACT 331

RESULT 11
BM792342/c
LOCUS BM792342 684 bp mRNA linear EST 05-MAR-2002
DEFINITION K-EST0072545 S22SNU16n1 Homo sapiens cDNA clone S22SNU16n1-13-E09
5', mRNA sequence.
ACCESSION BM792342
VERSION BM792342.1 GI:19140574
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 684)
AUTHORS Kim N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Soeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 13 row: E column: 09
High quality sequence stop: 684.
FEATURES
    source
    1..684
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="S22SNU16n1-13-E09"
        /sex="F"
        /tissue_type="Ascites"
        /cell_type="Lymphoblast-like"
        /cell_line="SNU-16"
        /lab_host="DH10B"
        /clone_lib="S22SNU16n1"
        /note="Organ: Stomach; Vector: pT73-Pac; Site 1: EcoRI;
        Site 2: NotI; The S22SNU16 library was contributed by the
        Soares laboratory and it was constructed as described by
        Ronaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome
        Research 6(9): 791-806. RNA was prepared from harvested
        cells of SNU-16 culture. SNU-16 cell was obtained from
        Korean Cell Line Bank (KCLB). SNU-16 was established from
        ascitic fluids of Korean patients by Park J.G. et al.
        (1990), Cancer Res 50: 2773-2780."
ORIGIN
    Query Match 35.9%; Score 160.2; DB 4; Length 684;
    Best Local Similarity 84.9%; Pred. No. 7.7e-18;
    Matches 191; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGATGCCACCGCCGGCTAA 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 439 CTCCTGCCTCACCCTCCCGAGTAGCTGGGATTACAGGATGCCACCGCCGGCCAA 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTGGTTCAGGCTGGTCTCGAACTTC 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 TTTTGTAT-TTTTAGTAGAGACGGCGTTCTCCATGTGGTCAAGCTGGTCTCCAACCTCC 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 AAACCTCAGGTGATCCGCCCTCGGCTCCCAAGTGTAGGATTACAGCGGTGAGCC 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 CAACTCAGGTGATCCACTCGGCTCCCAAGTGTGGGATTACAGCGGTGTGCC 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 203 ACCGGCTCAGCTGGGACACCTTTTCTTACATCTTCAAGTGCT 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 203 ACCGGCTCAGCTGGGACACCTTTTCTTACATCTTCAAGTGCT 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 ACCAGCCAGCGCTGGGTGTTTCTTTCAGCTCTCCAGTACT 108

RESULT 10
AW575808
LOCUS AW575808 566 bp mRNA linear EST 15-MAR-2000
DEFINITION UI-HF-BNO-aff-g-02-0-UT.s1 NIH_MGC_50 Homo sapiens cDNA clone
IMAGE:3066794 3', mRNA sequence.
ACCESSION AW575808
VERSION AW575808.1 GI:7247347
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 566)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a Bonafide poly A
tail. Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
DNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bbrp/image/image.html
The following repetitive elements were found in this cDNA
sequence:
10-299, >ALU
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
    source
    1..566
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:3066794"
        /tissue_type="lymph"
        /cell_type="germinal center B cells"
        /cell_line="MGC85"
        /lab_host="DH10B (LTI)"
        /clone_lib="NIH_MGC 50"
        /notes="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI;
        Constructed from size fractionated cytoplasmic mRNA
        (3.5-4.4kb). Directionally cloned. Cells provided by
        Louis M. Staudt, Ph.D. Library preparation by Maria de
        Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
ORIGIN
    Query Match 35.9%; Score 160.2; DB 2; Length 566;
    Best Local Similarity 84.9%; Pred. No. 8.1e-18;
    Matches 191; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY 23 CTGCAACCTCAGCCTCCCGAGTAGCTGGGATTACAGGATGCCACCGCCGGCTAA 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 CTCCTGCCTCACCCTCCCGAGTAGCTGGGATTACAGGATGCCACCGCCGGCCAA 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 TTTTGTATCTTTTAGTAGAGACGGCGTTCTCCATGTGGTTCAGGCTGGTCTCGAACTTC 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168 TTTTGTAT-TTTTAGTAGAGACGGCGTTCTCCATGTGGTCAAGCTGGTCTCCAACCTCC 226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 AAACCTCAGGTGATCCGCCCTCGGCTCCCAAGTGTAGGATTACAGCGGTGAGCC 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 227 CAACTCAGGTGATCCACTCGGCTCCCAAGTGTGGGATTACAGCGGTGTGCC 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 906)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAMI0389 row: d column: 19  
High quality sequence start: 4  
High quality sequence stop: 674.  
Location/Qualifiers  
1..906  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4509282"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_92"  
/notes="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 35.9%; Score 160.2; DB 4; Length 906;  
Best Local Similarity 84.9%; Pred. No. 7.2e-18;  
Matches 191; Conservative 0; Mismatches 33; Indels 1; Gaps 1;  
QY 23 CTCGAACCTCAGGCTCCCGAGTACGCGGCTCCCAAGTGTCTAGGATTCAGGCGGTGAGCC 82  
Db 482 CTCCTCGCTCACCTCCCGAGTACGCGGCTCCCAAGTGTCTAGGATTCAGGCGGTGAGCC 423  
QY 83 TTTTGTATCTTTAGTAGAGACGCGCTCTCCATGTGTGTAGGCTGGTCTCGAACTTC 142  
Db 422 TTTTGTAT- TTTTAGTAGAGACGCGGTTCTCCATGTGTGTAGGCTGGTCTCGAACTTC 364  
QY 143 AAACCTCAGGTATCCGCGCGCTCCGCGCTCCCAAGTGTCTAGGATTCAGGCGGTGAGCC 202  
Db 363 CAACCTCAGGTATCACCTCGCTCCGCGCTCCCAAGTGTCTAGGATTCAGGCGGTGAGCC 304  
QY 203 ACCGCGCTCAGCTGGGAACACCTTTTCTTACATCTTCAAGTGTCT 247  
Db 303 ACCACGCCAGCCTTGGGTGTTTTTCTTTTCAGTCTCTCCAGTACT 259

RESULT 14  
CN426512/c  
LOCUS 714 bp mRNA linear EST 16-MAY-2004  
DEFINITION 17000470665591 GRN\_ES Homo sapiens CDNA 5', mRNA sequence.  
ACCESSION CN426512  
VERSION CN426512.1 GI:47414106  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 714)  
REFERENCE Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,  
AUTHORS Li,Y., Xu,C., Pang,R., Guegler,K., Rao,M.S., Mandalam,R.,  
Lebkowski,J and Stanton,L.W.

